European Journal of Applied Mathematics (2025), 1–33
doi:10.1017/S0956792524000901

CAMBRIDGE UNIVERSITY PRESS

PAPER

On a general matrix-valued unbalanced optimal transport problem

Bowen Li[1] and Jun Zou[2]

[1]Department of Mathematics, City University of Hong Kong, Kowloon Tong, Hong Kong, China
[2]Department of Mathematics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
Corresponding author: Bowen Li; Email: bowen.li@cityu.edu.hk

Received: 22 October 2023; **Revised:** 27 June 2024; **Accepted:** 31 December 2024

Keywords: Unbalanced optimal transport; Matrix-valued measure; Geodesic space; Metric cone

2020 Mathematics Subject Classification: 47A56, 49Q20 (Primary); 28A33, 51F99, 58B20 (Secondary)

Abstract
We introduce a general class of transport distances WB_Λ over the space of positive semi-definite matrix-valued Radon measures $\mathcal{M}(\Omega, \mathbb{S}^n_+)$, called the weighted Wasserstein–Bures distance. Such a distance is defined via a generalised Benamou–Brenier formulation with a weighted action functional and an abstract matricial continuity equation, which leads to a convex optimisation problem. Some recently proposed models, including the Kantorovich–Bures distance and the Wasserstein–Fisher–Rao distance, can naturally fit into ours. We give a complete characterisation of the minimiser and explore the topological and geometrical properties of the space $(\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda)$. In particular, we show that $(\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda)$ is a complete geodesic space and exhibits a conic structure.

1. Introduction

1.1. Classical optimal transport

Optimal transport (OT) [87, 89, 90] provides a versatile framework for defining metrics and studying geometric structures on probability measures. It has been an active research area over the past decades with fruitful applications in various areas, including functional inequalities [68, 76, 88], gradient flow [51, 75], and more recently, image processing and machine learning [3, 37, 43]. The OT problem was first proposed by Monge in 1781 [72]: given probabilities ρ_0 and ρ_1, find a measure-preserving transport map T minimising

$$\min_{T_\# \rho_0 = \rho_1} \int |x - T(x)|^2 \, \mathrm{d}\rho_0(x) \,. \tag{1.1}$$

However, its solution (i.e., the OT map) may not exist. This question remained open for a long time until 1942 when Kantorovich introduced a relaxed problem based on the so-called transport plans [52]:

$$\mathrm{W}_2^2(\rho_0, \rho_1) := \min \left\{ \int |x - y|^2 \, \mathrm{d}\gamma \,; \gamma \text{ is a probability with } (\pi^x_\# \gamma, \pi^y_\# \gamma) = (\rho_0, \rho_1) \right\}, \tag{1.2}$$

where $\pi^x_\# \gamma$ and $\pi^y_\# \gamma$ are the first and second marginals of γ, respectively. The 2-Wasserstein distance (1.2) turns out to exhibit intriguing mathematical properties. Brenier [14] proved that under mild conditions, the OT map T to (1.1) exists and is uniquely given by the gradient of a convex function φ. Thanks to the measure-preserving property of the transport map $T = \nabla \varphi$, it is easy to see that φ satisfies the Monge–Ampère equation, which provides a PDE-based approach for solving the OT problem (1.1). One can also show that $(\mathrm{id}, \nabla \varphi)_\# \rho_0$ gives a minimiser to (1.2). Equipped with the distance $\mathrm{W}_2(\cdot, \cdot)$, the

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

probability measure space becomes a geodesic space, where the geodesic is characterised by McCann's displacement interpolation $\rho_t := ((1-t)I + t\nabla\varphi)_{\#}\rho_0$ [71]. In Benamou and Brenier's seminal work [8], an equivalent fluid mechanics formulation was proposed for computational purposes:

$$\mathrm{W}_2^2(\rho_0, \rho_1) = \min_{\rho,m} \left\{ \frac{1}{2} \iint \rho^{-1}|m|^2 \,\mathrm{d}t\,\mathrm{d}x \,;\, \partial_t \rho + \operatorname{div} m = 0 \right\} . \qquad (\mathcal{P}_{\mathrm{W}_2})$$

This dynamic point of view has since stimulated numerous follow-up studies, including the present work. We refer the interested readers to [89, 90] for the precise statements of aforementioned results and a detailed overview.

1.2. Unbalanced optimal transport

Although the OT theory has become a popular tool in learning theory and data science for its geometric nature and capacity for large-scale simulation, a limitation is that the associated metric is only defined for measures of equal mass, while in many applications, it is more desirable to allow measures with different masses. This leads to the problem of extending the classical OT theory to the unbalanced case. The early effort in this direction may date back to the works [53, 54] by Kantorovich and Rubinshtein in the 1950s, where a simple static formulation with an extended Kantorovich norm was introduced. The underlying idea is to allow the mass to be sent to (or come from) a point at infinity, which was further investigated and extended in [49, 50]. Similarly, Figalli and Gigli [39] introduced an unbalanced transportation distance via a variant of Kantorovich formulation (1.2) by allowing taking the mass from (or giving it back to) the boundary of the domain. Another closely related approach is the optimal partial transport [20, 38], which is also based on (1.2) but involves a relaxed constraint $(\pi^x_{\#}\gamma, \pi^y_{\#}\gamma) \le (\rho_0, \rho_1)$ and a shifted cost $|x-y|^2 - \alpha$.

In addition to the static models, there is a large number of works devoted to defining an unbalanced OT model via a dynamic formulation in the spirit of [8]; see for example [7, 27, 66, 69, 79]. In these works, a source term and a corresponding penalisation term are introduced in the continuity equation and the action functional, respectively, in order to model the mass change. In particular, Piccoli and Rossi [78, 79] defined a generalised Wasserstein distance by relaxing the marginal constraint $(\pi^x_{\#}\gamma, \pi^y_{\#}\gamma) = (\rho_0, \rho_1)$ by a total variation regularisation, which turns out to be equivalent to the optimal partial transport in certain scenarios [27]. Moreover, an equivalent dynamic formulation has also been given in ref. [79]. Later, a new transport model, called the Wasserstein–Fisher–Rao (WFR) or Hellinger–Kantorovich distance (in this work we adopt the former one), was introduced independently and almost simultaneously by three research groups with different perspectives and techniques [27, 56, 64]. This model can be regarded as an inf-convolution of the Wasserstein and Fisher–Rao metric tensors, as the name suggests. In their subsequent work [29], Chizat et al. presented a class of unbalanced transport distances in a unified framework via both static and dynamic formulations, thanks to the notions of semi-couplings and Lagrangians. Meanwhile, Liero et al. [65] proposed a related optimal entropy-transport approach and discussed its properties in detail. It was proved that both the optimal partial transport and the WFR distance can be viewed as the special cases of the general frameworks in refs. [29, 65]. After that, the unbalanced OT theory is further developed in various directions, such as gradient flows [57, 59], Sobolev inequalities [58] and the JKO scheme [41, 44]. We also want to mention a recent work [67] by Lombardini and Rossi, which gave a negative answer to an interesting question of whether it is possible to define an unbalanced transport distance that coincides with the Wasserstein one when the measures are of equal mass.

1.3. Noncommutative optimal transport

More recently, there is also an increasing interest in generalising the OT theory to the noncommutative setting, namely, the quantum states or matrix-valued measures. The first line of research is motivated by the ergodicity of open quantum dynamics [48, 55, 74]. In the seminal works [21, 22] by Carlen and Maas, a quantum Wasserstein distance was introduced with a Benamou–Brenier dynamic formulation such that a primitive quantum Markov semigroup satisfying the detailed balance condition can be formulated as the gradient flow of the logarithmic relative entropy, which opens the door to investigating the noncommutative functional inequalities via the gradient flow techniques and the geodesic convexity; see for example [31, 62, 84, 93]. Meanwhile, Golse et al. proposed another quantum transport model via a generalised Monge–Kantorovich formulation, when they studied the mean-field and classical limits of the Schrödinger equation; see [45–47]. Other static quantum Wasserstein distances can be found in refs. [30, 32, 33], just to name a few.

The second research line is driven by the advances in diffusion tensor imaging [61, 92], where a tensor field (usually, a positive semi-definite matrix) is generated at each spatial position to encode the local diffusivity of water molecules in the brain. It gives rise to a natural question of how to compare two brain tensor fields, or mathematically how to define a reasonable distance between matrix-valued measures. Chen et al. [23, 24] introduced a dynamic matricial Wasserstein distance for matrix-valued densities with unit mass, drawing inspiration from ref. [8] and leveraging the Lindblad equation in quantum mechanics, which was later extended to the unbalanced case [25] in a manner similar to [27]. In particular, Brenier and Vorotnikov [16] recently proposed a different dynamic OT model for unbalanced matrix-valued measures called the Kantorovich–Bures metric, which is motivated by the observation in ref. [15] that the incompressible Euler equation admits a dual concave maximisation problem. Regarding static formulations, Peyré et al. [77] introduced a quantum transport distance with entropic regularisation inspired by [65] and proposed an associated scaling algorithm that generalised the results in ref. [28]. Additionally, Ryu et al. defined a matrix OT model of order 1 by a Beckmann-type flux formulation and presented a scalable and parallelisable numerical method. Applications in tensor field imaging were also explored in ref. [77, 86].

1.4. Contribution

The initial motivation for this work is the numerical study of the unbalanced matricial OT models proposed in ref. [16, 25]; see ($\mathcal{P}_{\mathrm{WB}}$) and ($\mathcal{P}_{2,\mathrm{FR}}$). We find that despite their distinct formulations, these models actually share many mathematical properties. In this work, we consider an abstract continuity equation $\partial_t \mathsf{G} + \mathsf{D}\mathsf{q} = \mathsf{R}^{\mathrm{sym}}$ in Definition 3.4 with D being a first-order constant coefficient linear differential operator such that $\mathsf{D}^*(I) = 0$, in analogy with the one $\partial_t \mathsf{G} + 2(\mathsf{L}^* \circ \mathsf{P})\,\mathsf{q} = 0$ for the matrix-valued optimal ballistic transport problem (cf. [91, (1.4)–(1.5)]). Here, $\mathsf{q}(t, x)$ can be intuitively seen as a momentum variable; $\mathsf{D}q$ is the matricial analogue of the advection term $\operatorname{div} m$ in ($\mathcal{P}_{\mathrm{W}_2}$) controlling the mass transportation in space and between components; $\mathsf{R}^{\mathrm{sym}}$ is the reaction part describing the variation of mass. Then, thanks to the weighted infinitesimal cost $J_\Lambda(G_t, q_t, R_t) = \frac{1}{2}(q_t\Lambda_1^\dagger)\cdot G_t^\dagger(q_t\Lambda_1^\dagger) + \frac{1}{2}(R_t\Lambda_2^\dagger)\cdot G_t^\dagger(R_t\Lambda_2^\dagger)$ given in Proposition 3.1 with the weight matrices Λ_1 and Λ_2 representing the contributions of each component of q and G in J_Λ, we define a general matrix-valued unbalanced OT distance $\mathrm{WB}_\Lambda(\cdot,\cdot)$ ($\mathcal{P}$) as a convex optimisation, similarly to the classical case ($\mathcal{P}_{\mathrm{W}_2}$), which we call the weighted Wasserstein–-Bures distance; see Definition 3.8. We note that the problems ($\mathcal{P}_{\mathrm{WB}}$) and ($\mathcal{P}_{2,\mathrm{FR}}$), as well as the scalar WFR distance ($\mathcal{P}_{\mathrm{WFR}}$), can be viewed as the special instances of our model ($\mathcal{P}$). See Section 7 for more details.

Our main contribution is a comprehensive and self-contained study of the properties of the weighted distance WB_Λ on the positive semi-definite matrix-valued Radon measure space $\mathcal{M}(\Omega, \mathbb{S}_+^n)$. We establish the a priori estimates for solutions of the abstract continuity equation (3.13) in Lemmas 3.9, 3.12 and Proposition 3.13, which consequently gives the well-posedness of the model ($\mathcal{P}$) and a useful compactness result (Proposition 3.18). Then, by leveraging tools from convex analysis, we show the existence of the minimiser (i.e., the minimising geodesic) to ($\mathcal{P}$) with a characterisation of the optimality conditions; see Theorems 4.2 and 4.5. Moreover, we prove that the topology induced by $\mathrm{WB}_\Lambda(\cdot,\cdot)$ is stronger than

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

the weak* one, and study the limit model when a weight matrix goes to zero; see Propositions 5.2 and 4.6, respectively. With the help of these results, in Theorem 5.5 and Corollary 5.7, we characterise the absolutely continuous curve with respect to the metric WB_Λ and show that $(\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda)$ is a complete geodesic space. We further consider its conic structure and prove in Theorem 6.3 that the space $(\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda)$ is a metric cone over $(\mathcal{M}_1, \mathrm{SWB}_\Lambda)$, where $\mathcal{M}_1$ is a normalised matrix-valued measure space (6.2), which corresponds to a noncommutative probability space, and SWB_Λ is the spherical distance (6.1) induced by WB_Λ. Recalling the Riemannian interpretation in Corollary 5.8, we can formally view $(\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda)$ as a Riemannian manifold and $\mathcal{M}_1$ as its submanifold with the induced metric SWB_Λ, which allows developing the Otto calculus in the spirit of [76]. These results can be readily applied to the models ($\mathcal{P}_{\mathrm{WB}}$) and ($\mathcal{P}_{2,\mathrm{FR}}$), which lay a solid mathematical foundation for the distance ($\mathcal{P}_{2,\mathrm{FR}}$) and complement the results in ref. [16] for ($\mathcal{P}_{\mathrm{WB}}$) (note that our approach is quite different from theirs).

In the companion work [63], we have designed a convergent discretisation scheme for the general model ($\mathcal{P}$), which directly applies to the Kantorovich–Bures distance ($\mathcal{P}_{\mathrm{WB}}$) [16], the matricial interpolation distance ($\mathcal{P}_{2,\mathrm{FR}}$) [25] and the WFR metric ($\mathcal{P}_{\mathrm{WFR}}$) [27], thanks to the discussion in Section 7 of the present work.

1.5. Layout

The rest of this work is organised as follows. In Section 2, we give a list of basic notations that will be used throughout this work and recall some preliminary results. In Section 3, we define a class of weighted Wasserstein–Bures distances for matrix-valued measures via a dynamic formulation. Sections 4 and 5 are devoted to its topological, metric and geometric properties, while in Section 6, we discuss its conic structure. In Section 7, we connect our general model with several existing models in the literature. Some auxiliary proofs are included in Appendix A.

2. Preliminaries and notation

2.1. Notation and convention

- We denote by $\mathbb{R}^{n\times m}$ the space of $n \times m$ real matrices. If $m = n$, we simply write it as $\mathbb{M}^n$. Moreover, we use $\mathbb{S}^n$, $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ to denote symmetric matrices, positive semi-definite matrices and positive definite matrices, respectively. $\mathbb{A}^n$ denotes the space of $n \times n$ antisymmetric matrices.
- We denote by $|\cdot|$ the Euclidean norm on $\mathbb{R}^n$. We equip the matrix space $\mathbb{R}^{n\times m}$ with the Frobenius inner product $A \cdot B = Tr(A^{\mathrm{T}}B)$ and the associated norm $\|A\|_{\mathrm{F}} = \sqrt{A \cdot A}$.
- The symmetric and antisymmetric parts of $A \in \mathbb{M}^n$ are given by
$$A^{\mathrm{sym}} = (A + A^{\mathrm{T}})/2\,, \quad A^{\mathrm{ant}} = (A - A^{\mathrm{T}})/2\,, \tag{2.1}$$
respectively. We also write $A \preceq B$ (resp., $A \prec B$) for $A, B \in \mathbb{S}^n$ if $B - A \in \mathbb{S}^n_+$ (resp., $B - A \in \mathbb{S}^n_{++}$).
- $\mathcal{X}$ denotes a generic compact separable metric space with Borel σ-algebra $\mathscr{B}(\mathcal{X})$, unless otherwise specified.
- $C(\mathcal{X}, \mathbb{R}^n)$ denotes the space of $\mathbb{R}^n$-valued continuous functions on $\mathcal{X}$ with the supremum norm $\|\cdot\|_\infty$. Its dual space, denoted by $\mathcal{M}(\mathcal{X}, \mathbb{R}^n)$, is $\mathbb{R}^n$-valued Radon measure space with the total variation norm $\|\cdot\|_{\mathrm{TV}}$.
- Let $\mathcal{B}$ be a Banach space with the dual space $\mathcal{B}^*$. We denote by $\langle \cdot, \cdot \rangle_{\mathcal{B}}$ the duality pairing between $\mathcal{B}$ and $\mathcal{B}^*$. When $\mathcal{B} = C(\mathcal{X}, \mathbb{R}^n)$, we usually write it as $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ for short. We will also consider the weak and weak* convergences on $\mathcal{B}$ and $\mathcal{B}^*$, respectively. In particular, a sequence of measures $\{\mu_j\}$ weak* converges to $\mu \in \mathcal{M}(\mathcal{X}, \mathbb{R}^n)$ if for any $\phi \in C(\mathcal{X}, \mathbb{R}^n)$, there holds $\langle \mu_j, \phi \rangle_{\mathcal{X}} \to \langle \mu, \phi \rangle_{\mathcal{X}}$ as $j \to +\infty$.

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

- Let $\mathbb{R}_+ := [0,\infty)$, and $\mathcal{M}(\mathcal{X},\mathbb{R}_+)$ be the space of nonnegative finite Radon measures. For $\mu \in \mathcal{M}(\mathcal{X},\mathbb{R}^n)$, we have an associated variation measure $|\mu| \in \mathcal{M}(\mathcal{X},\mathbb{R}_+)$ such that $\mathrm{d}\mu = \sigma \mathrm{d}|\mu|$ with $|\sigma(x)| = 1$ for $|\mu|$-a.e. $x \in \mathcal{X}$, where $\sigma : \mathcal{X} \to \mathbb{R}^n$ is the Radon–Nikodym derivative (density) of μ with respect to $|\mu|$ [36, 85].
- We identify the space of matrix-valued Radon measures $\mathcal{M}(\mathcal{X},\mathbb{R}^{n\times m})$ with $\mathcal{M}(\mathcal{X},\mathbb{R}^{nm})$ by vectorisation. It is easy to see that both sets of $\mathbb{S}^n$-valued Radon measures $\mathcal{M}(\mathcal{X},\mathbb{S}^n)$ and $\mathbb{S}^n_+$-valued Radon measures $\mathcal{M}(\mathcal{X},\mathbb{S}^n_+)$ are closed in $\mathcal{M}(\mathcal{X},\mathbb{M}^n)$ with respect to the weak* topology [35, Theorem 3.5]. Moreover, we have the following characterisation:
$$(C(\mathcal{X},\mathbb{S}^n))^* \simeq (C(\mathcal{X},\mathbb{M}^n)/C(\mathcal{X},\mathbb{A}^n))^* \simeq \mathcal{M}(\mathcal{X},\mathbb{S}^n)\,,$$
where $\simeq$ means the isometric isomorphism and $C(\mathcal{X},\mathbb{M}^n)/C(\mathcal{X},\mathbb{A}^n)$ is the quotient space. Indeed, we observe that $\mu \in \mathcal{M}(\mathcal{X},\mathbb{S}^n) \subset \mathcal{M}(\mathcal{X},\mathbb{M}^n) \simeq C(\mathcal{X},\mathbb{M}^n)^*$ if and only if its induced linear functional on $C(\mathcal{X},\mathbb{M}^n)$ has the kernel $C(\mathcal{X},\mathbb{A}^n)$, which yields, by [17, Proposition 11.9],
$$(C(\mathcal{X},\mathbb{M}^n)/C(\mathcal{X},\mathbb{A}^n))^* \simeq \mathcal{M}(\mathcal{X},\mathbb{S}^n)\,.$$
Meanwhile, $C(\mathcal{X},\mathbb{S}^n) \simeq C(\mathcal{X},\mathbb{M}^n)/C(\mathcal{X},\mathbb{A}^n)$ is a consequence of $\mathbb{S}^n \perp \mathbb{A}^n$ and $\mathbb{S}^n \simeq \mathbb{M}^n/\mathbb{A}^n$.
- For $\mu \in \mathcal{M}(\mathcal{X},\mathbb{S}^n_+)$, we define an associated trace measure $Tr\mu$ by the set function $E \to Tr(\mu(E))$, $E \in \mathscr{B}(\mathcal{X})$. It is clear that $0 \preceq \mu(E) \preceq Tr(\mu(E))I$ and $Tr\mu$ is equivalent to $|\mu|$, denoted by $Tr\mu \sim |\mu|$. That is,
$$|\mu| \ll Tr\mu \quad \text{and} \quad Tr\mu \ll |\mu|\,. \tag{2.2}$$
We will usually use $Tr\mu$ as the dominant measure for $\mu \in \mathcal{M}(\mathcal{X},\mathbb{S}^n_+)$. In addition, note that for $\lambda \in \mathcal{M}(\mathcal{X},\mathbb{R}_+)$ with $|\mu| \ll \lambda$, there holds $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \in \mathbb{S}^n_+$ for λ-a.e. $x \in \mathcal{X}$, which is an equivalent characterisation of $\mathcal{M}(\mathcal{X},\mathbb{S}^n_+)$.
- We will use sans serif letterforms to denote vector-valued or matrix-valued measures, e.g., $\mathsf{A} \in \mathcal{M}(\mathcal{X},\mathbb{M}^n)$, while letters with serifs are reserved for their densities with respect to some reference measure, e.g., $A_\lambda := \frac{\mathrm{d}\mathsf{A}}{\mathrm{d}\lambda}$ for $|\mathsf{A}| \ll \lambda$. The symmetric and antisymmetric parts $\mathsf{A}^{\mathrm{sym}}$ and $\mathsf{A}^{\mathrm{ant}}$ of $\mathsf{A} \in \mathcal{M}(\mathcal{X},\mathbb{M}^n)$ are defined as in (2.1).
- We identify a measure and its density with respect to the Lebesgue measure (if exists) unless otherwise specified.
- For $\lambda \in \mathcal{M}(\mathcal{X},\mathbb{R}_+)$, we denote by $L^p_\lambda(\mathcal{X},\mathbb{R}^n)$ with $p \in [1,+\infty]$ the standard space of p-integrable $\mathbb{R}^n$-valued functions. For $\mathsf{G} \in \mathcal{M}(\mathcal{X},\mathbb{S}^n_+)$, we consider the space of $\mathbb{R}^{n\times m}$-valued measurable functions endowed with the semi-inner product:
$$\langle P, Q\rangle_{L^2_{\mathsf{G}}(\mathcal{X})} := \langle \mathsf{G}, QP^{\mathrm{T}}\rangle_{\mathcal{X}} = \int_{\mathcal{X}} P\cdot(\mathrm{d}\mathsf{G}\,Q) = \int_{\mathcal{X}} P\cdot\big(G_\lambda Q\big)\,\mathrm{d}\lambda\,, \tag{2.3}$$
where λ is a reference measure such that $|\mathsf{G}| \ll \lambda$ and G_λ is the density. Noting that $\|Q\|_{L^2_{\mathsf{G}}(\mathcal{X})} = 0$ is equivalent to $G_\lambda Q = 0$ for λ-a.e. $x \in \mathcal{X}$, the kernel of the seminorm $\|\cdot\|_{L^2_{\mathsf{G}}(\mathcal{X})}$ is given by $\{Q\,;\,\mathrm{Ran}(Q) \in \mathrm{Ker}(G_\lambda)\,,\ \lambda\text{-a.e.}\}$. Then, we define the Hilbert space $L^2_{\mathsf{G}}(\mathcal{X},\mathbb{R}^{n\times m})$ as the quotient space by $\mathrm{Ker}\big(\|\cdot\|_{L^2_{\mathsf{G}}(\mathcal{X})}\big)$.

2.2. Preliminaries

We denote by $A^\dagger \in \mathbb{R}^{m\times n}$ the pseudoinverse of a matrix $A \in \mathbb{R}^{n\times m}$. If $A \in \mathbb{S}^n$ has the eigendecomposition $A = O\Sigma O^{\mathrm{T}}$, then $A^\dagger = O\Sigma^\dagger O^{\mathrm{T}}$ with $\Sigma^\dagger = \mathrm{diag}(\lambda_1^{-1},\ldots,\lambda_s^{-1},0,\ldots,0)$, where O is an orthogonal matrix and $\Sigma = \mathrm{diag}(\lambda_1,\ldots,\lambda_s,0,\ldots,0)$ is a diagonal matrix with $\{\lambda_i\}$ being nonzero eigenvalues of A.

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

Lemma 2.1. *The following properties hold:*

1. *If* $A \succeq B \succeq 0$ *and* $\mathrm{Ran}(A) = \mathrm{Ran}(B)$*, then* $B^\dagger \succeq A^\dagger$.
2. *The cone* $\mathbb{S}^n_+$ *in* $\mathbb{S}^n$ *is self-dual, that is,* $(\mathbb{S}^n_+)^* := \{B \in \mathbb{S}^n\,;\, Tr(AB) \ge 0\,,\ \forall A \in \mathbb{S}^n_+\} = \mathbb{S}^n_+$.
3. *If* $A, B \succeq 0$ *and* $A \cdot B = 0$*, then* $\mathrm{Ran}B \subset \mathrm{Ker}A$*, equivalently,* $\mathrm{Ran}A \subset \mathrm{Ker}B$.
4. *For* $A \in \mathbb{S}^n_+, M \in \mathbb{R}^{n\times m}$*, there holds*

$$(AM)\cdot M \le Tr(A)\|M\|^2_{\mathrm{F}}\,. \tag{2.4}$$

Remark 2.2. The range condition $\mathrm{Ran}(A) = \mathrm{Ran}(B)$ for the first statement in Lemma 2.1 above is necessary, due to the example $A = \mathrm{diag}(1, 1, 1, 0)$ and $B = \mathrm{diag}(1, 1, 0, 0)$. Moreover, we remark that for $\mathsf{G} \in \mathcal{M}(\mathcal{X}, \mathbb{S}^n_+)$, there holds $L^2_{Tr\mathsf{G}}(\mathcal{X}, \mathbb{R}^n) \subset L^2_{\mathsf{G}}(\mathcal{X}, \mathbb{R}^n)$ by (2.4), while the converse is not true; see [35] for the counterexample.

Proof. We only prove the first statement, as the others are direct. We first note that the orthogonal projection onto $\mathrm{Ran}(A) = \mathrm{Ran}(B)$ is given by $\mathbb{P} = \sqrt{B}^\dagger B \sqrt{B}^\dagger = \sqrt{A}^\dagger A \sqrt{A}^\dagger$. By $A - B \succeq 0$, we have $\sqrt{B}^\dagger A \sqrt{B}^\dagger - \mathbb{P} \succeq 0$, which means that all the eigenvalues of the matrix $\sqrt{B}^\dagger A \sqrt{B}^\dagger$ restricted on its invariant subspace $\mathrm{Ran}(A) = \mathrm{Ran}(B)$ is greater than or equal to one. It is easy to see that $\sqrt{B}^\dagger A \sqrt{B}^\dagger$ and $\sqrt{A} B^\dagger \sqrt{A}$ have the same eigenvalues. Hence, we find $\sqrt{A} B^\dagger \sqrt{A} - \mathbb{P} \succeq 0$, which gives $B^\dagger \succeq A^\dagger$ by conjugating with $\sqrt{A}^\dagger$. □

The next lemma is about the measurability of matrix-valued functions.

Lemma 2.3. *Let* $A(x)$ *be a* $\mathbb{S}^n$*-valued Borel measurable function on* $\mathcal{X}$*. Then, it holds that*

1. *The eigenvalues* $\{\lambda_{A,i}(x)\}_{i=1}^n$ *of* $A(x)$ *in nondecreasing order are measurable, and the corresponding eigenvectors* $\{u_{A,i}(x)\}_{i=1}^n$ *can also be selected to be measurable and form an orthonormal basis of* $\mathbb{R}^n$ *for every* $x \in \mathcal{X}$.
2. *The pseudoinverse* $A^\dagger(x)$ *of* $A(x)$ *is measurable, and the square root* $A^{1/2}(x)$ *of* $A(x) \in \mathbb{S}^n_+$ *is measurable.*

The first and second properties are from [81] and [82] with the continuity of $A^{1/2}$ in $A \in \mathbb{S}^n_+$, respectively. In fact, Powers–Størmer inequality [80] gives

$$\left\|\sqrt{A} - \sqrt{B}\,\right\|^2_{\mathrm{F}} \le \sqrt{n}\|A - B\|_{\mathrm{F}}\,, \quad \forall A, B \in \mathbb{S}^n_+\,. \tag{2.5}$$

We finally recall some concepts and useful results from convex analysis. Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be an extended real-valued function on a Banach space X. We denote by $\partial f(x)$ its subgradient at $x \in X$ and by $dom(f) := f^{-1}(\mathbb{R})$ its domain. We say that f is proper if $dom(f) \neq \varnothing$; and that f is positively homogeneous of degree k if for all $x \in X$ and $\alpha > 0$, $f(\alpha x) = \alpha^k f(x)$. The conjugate function f^* of f is defined by

$$f^*(x^*) = \sup_{x\in X}\langle x^*, x\rangle_X - f(x)\,, \quad \forall x^* \in X^*\,, \tag{2.6}$$

which is convex and lower semicontinuous with respect to the weak* topology of X^*. The following two lemmas are from [4, Proposition 2.33] and [12, Proposition 2.5], respectively.

Lemma 2.4 (Subgradient). *Let* $f : X \to \mathbb{R} \cup \{+\infty\}$ *be a proper convex function on a Banach space* X*. Then, the following three properties are equivalent: (i)* $x^* \in \partial f(x)$*; (ii)* $f(x) + f^*(x^*) = \langle x^*, x\rangle_X$*; (iii)* $f(x) + f^*(x^*) \le \langle x^*, x\rangle_X$*. In addition, if* f *is lower semicontinuous, then all of these properties are equivalent to* $x \in \partial f^*(x^*)$.

Lemma 2.5 (Fenchel–Rockafellar duality). *Let* X *and* Y *be two Banach spaces and* $L : X \to Y$ *be a bounded linear operator with the adjoint* $L^* : Y^* \to X^*$*. Let* f *and* g *be two proper lower semicontinuous*

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

convex functions defined on X and Y valued in $\mathbb{R} \cup \{+\infty\}$, respectively. If there exists $x \in dom(f)$ such that g is continuous at Lx, then

$$\sup_{x\in X} -f(-x) - g(Lx) = \inf_{y^*\in Y^*} f^*(L^*y^*) + g^*(y^*)\,, \tag{2.7}$$

and the inf *in* (2.7) *can be attained. Moreover, the* sup *in* (2.7) *is attained at $x \in X$ if and only if there exists a $y^* \in Y^*$ such that $Lx \in \partial g^*(y^*)$ and $L^*y^* \in \partial f(-x)$, in which case y^* also achieves the* inf *in* (2.7).

3. Definition and basic properties

We shall introduce a new family of distances on the matrix-valued Radon measure space $\mathcal{M}(\Omega, \mathbb{S}^n_+)$ based on a dynamic OT formulation, which will be the central object of this work.

3.1. Action functional

To define our dynamic OT model over the space of $\mathbb{S}^n_+$-valued measures, the starting point is a weighted action functional. Let $n, k, m \in \mathbb{N}$ be positive integers and $\Lambda := (\Lambda_1, \Lambda_2)$ be a pair of matrices with $\Lambda_1 \in \mathbb{S}^k_+$ and $\Lambda_2 \in \mathbb{S}^m_+$. We define the following closed convex set:

$$\mathcal{O}_\Lambda = \left\{(A, B, C) \in \mathbb{S}^n \times \mathbb{R}^{n\times k} \times \mathbb{R}^{n\times m}\,;\, A + \frac{1}{2}B\Lambda_1^2B^{\mathrm{T}} + \frac{1}{2}C\Lambda_2^2C^{\mathrm{T}} \preceq 0\right\}. \tag{3.1}$$

Note that its characteristic function:

$$\iota_{\mathcal{O}_\Lambda} := \begin{cases} 0, & (A, B, C) \in \mathcal{O}_\Lambda\,, \\ +\infty, & (A, B, C) \notin \mathcal{O}_\Lambda\,, \end{cases}$$

is proper lower semicontinuous and convex [6, Lemma 1.24]. We denote by J_Λ the conjugate function (2.6) of $\iota_{\mathcal{O}_\Lambda}$ and derive the explicit expressions for J_Λ and its subgradient ∂J_Λ.

Proposition 3.1. *J_Λ is proper, positively homogeneous of degree one, lower semicontinuous and convex with the following representation:*

$$J_\Lambda(X, Y, Z) = \frac{1}{2}(Y\Lambda_1^\dagger)\cdot(X^\dagger Y\Lambda_1^\dagger) + \frac{1}{2}(Z\Lambda_2^\dagger)\cdot(X^\dagger Z\Lambda_2^\dagger)\,, \tag{3.2}$$

if $X \in \mathbb{S}^n_+$, $\mathrm{Ran}(Y^{\mathrm{T}}) \subset \mathrm{Ran}(\Lambda_1)$, $\mathrm{Ran}(Z^{\mathrm{T}}) \subset \mathrm{Ran}(\Lambda_2)$ *and* $\mathrm{Ran}([Y, Z]) \subset \mathrm{Ran}(X)$*; otherwise, $J_\Lambda(X, Y, Z) = +\infty$. Moreover, the subgradient of J_Λ at $(X, Y, Z) \in dom(J_\Lambda)$ is characterised by*

$$\partial J_\Lambda(X, Y, Z) = \left\{(A, B, C) \in \mathcal{O}_\Lambda\,;\, Y = XB\Lambda_1^2\,,\ Z = XC\Lambda_2^2\,,\ X\cdot\left(A + \frac{1}{2}B\Lambda_1^2B^{\mathrm{T}} + \frac{1}{2}C\Lambda_2^2C^{\mathrm{T}}\right) = 0\right\}. \tag{3.3}$$

$\partial J_\Lambda(X, Y, Z)$ is a singleton if and only if $(X, Y, Z) \in \mathbb{S}^n_{++} \times \mathbb{R}^{n\times k} \times \mathbb{R}^{n\times m}$ and $\Lambda_1 \in \mathbb{S}^k_{++}$, $\Lambda_2 \in \mathbb{S}^m_{++}$.

Proof. The properties of J_Λ are by [6, Proposition 14.11]. To derive the formula (3.2), by definition, we have

$$J_\Lambda(X, Y, Z) = \sup_{(A,B,C)\in\mathcal{O}_\Lambda} X\cdot A + Y\cdot B + Z\cdot C\,, \tag{3.4}$$

for $(X, Y, Z) \in \mathbb{S}^n \times \mathbb{R}^{n\times k} \times \mathbb{R}^{n\times m}$. We consider the following four cases.
Case I: $X \in \mathbb{S}^n\backslash\mathbb{S}^n_+$. We choose a vector $a \in \mathbb{R}^n$ such that $\langle a, Xa\rangle < 0$ and set $A = -\lambda aa^{\mathrm{T}} \preceq 0$ with $\lambda > 0$, $B = 0$ and $C = 0$ in (3.4). Then it follows that

$$J_\Lambda(X, Y, Z) \ge \sup_{\lambda>0} X\cdot(-\lambda aa^{\mathrm{T}}) = +\infty\,.$$

Case II: $\mathrm{Ran}(Y^{\mathrm{T}}) \not\subset \mathrm{Ran}(\Lambda_1)$ or $\mathrm{Ran}(Z^{\mathrm{T}}) \not\subset \mathrm{Ran}(\Lambda_2)$. It suffices to consider the case $\mathrm{Ran}(Y^{\mathrm{T}}) \not\subset \mathrm{Ran}(\Lambda_1)$, since the same argument applies to the other one. Without loss of generality, we let $Y = [y_1, \ldots, y_n]^{\mathrm{T}}$

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

with $y_i \in \mathbb{R}^k$ and $y_1 \notin \mathrm{Ran}(\Lambda_1)$. Thanks to $\Lambda_1 \in \mathbb{S}^k_+$, y_1 has the orthogonal decomposition:

$$y_1 = y_1^{(1)} + y_1^{(2)} \quad \text{with } y_1^{(1)} \in \mathrm{Ran}(\Lambda_1)\,,\ y_1^{(2)} \neq 0 \in \mathrm{Ker}(\Lambda_1)\,.$$

Taking $A = 0$, $B = \lambda\big[y_1^{(2)}, 0\big]^{\mathrm{T}}$ with $\lambda \in \mathbb{R}$ and $C = 0$ in (3.4), we have

$$J_\Lambda(X, Y, Z) \ge \sup_{\lambda > 0} \lambda \big|y_1^{(2)}\big|^2 = +\infty\,.$$

Case III: $\mathrm{Ran}([Y, Z]) \not\subset \mathrm{Ran}(X)$. It suffices to consider $\mathrm{Ran}(Y) \not\subset \mathrm{Ran}(X)$. We take (A, B, C) in (3.4) as:

$$A = -\frac{\lambda^2}{2}(\mathbb{P}_{\mathrm{Ker}(X)} Y \Lambda_1)(\mathbb{P}_{\mathrm{Ker}(X)} Y \Lambda_1)^{\mathrm{T}}\,,\ B = \lambda \mathbb{P}_{\mathrm{Ker}(X)} Y\,,\ C = 0\,,$$

with $\lambda > 0$, where $\mathbb{P}_{\mathrm{Ker}(X)} := I - X^\dagger X$ is the orthogonal projection onto $\mathrm{Ker}(X)$. A direct computation gives

$$\begin{aligned} J_\Lambda(X, Y, Z) &\ge \sup_{(A,B,0)\in \mathcal{O}_\Lambda} X \cdot A + Y \cdot B \\ &\ge \sup_{\lambda > 0} -\frac{\lambda^2}{2}(\mathbb{P}_{\mathrm{Ker}(X)} Y \Lambda_1) \cdot (X \mathbb{P}_{\mathrm{Ker}(X)} Y \Lambda_1) + \lambda Y \cdot (\mathbb{P}_{\mathrm{Ker}(X)} Y) \\ &\ge \sup_{\lambda > 0} \lambda (\mathbb{P}_{\mathrm{Ker}(X)} Y) \cdot (\mathbb{P}_{\mathrm{Ker}(X)} Y) = +\infty\,, \end{aligned}$$

since there holds $(\mathbb{P}_{\mathrm{Ker}(X)} Y) \cdot (\mathbb{P}_{\mathrm{Ker}(X)} Y) > 0$ by $\mathrm{Ran}(Y) \not\subset \mathrm{Ran}(X)$.

Case IV: $(X, Y, Z) \in \mathbb{S}^n_+ \times \mathbb{R}^{n\times k} \times \mathbb{R}^{n \times m}$ with $\mathrm{Ran}(Y^{\mathrm{T}}) \subset \mathrm{Ran}(\Lambda_1)$, $\mathrm{Ran}(Z^{\mathrm{T}}) \subset \mathrm{Ran}(\Lambda_2)$ and $\mathrm{Ran}([Y, Z]) \subset \mathrm{Ran}(X)$. For this case, we directly compute

$$X \cdot A + Y \cdot B + Z \cdot C = X \cdot \Big(A + \frac{1}{2} B \Lambda_1^2 B^{\mathrm{T}} + \frac{1}{2} C \Lambda_2^2 C^{\mathrm{T}}\Big) + Y \cdot B + Z \cdot C - X \cdot \Big(\frac{1}{2} B \Lambda_1^2 B^{\mathrm{T}} + \frac{1}{2} C \Lambda_2^2 C^{\mathrm{T}}\Big)\,, \tag{3.5}$$

and

$$\begin{aligned} Y \cdot B + Z \cdot C - \frac{1}{2} X \cdot \big(B \Lambda_1^2 B^{\mathrm{T}} + C \Lambda_2^2 C^{\mathrm{T}}\big) = &-\frac{1}{2}\Big\|\sqrt{X} B \Lambda_1 - \sqrt{X}^\dagger Y \Lambda_1^\dagger\Big\|_{\mathrm{F}}^2 - \frac{1}{2}\Big\|\sqrt{X} C \Lambda_2 - \sqrt{X}^\dagger Z \Lambda_2^\dagger\Big\|_{\mathrm{F}}^2 \\ &+ \frac{1}{2}\Big\|\sqrt{X}^\dagger Y \Lambda_1^\dagger\Big\|_{\mathrm{F}}^2 + \frac{1}{2}\Big\|\sqrt{X}^\dagger Z \Lambda_2^\dagger\Big\|_{\mathrm{F}}^2\,, \end{aligned} \tag{3.6}$$

where we have used

$$Y \cdot B + Z \cdot C = \big(\sqrt{X}\sqrt{X}^\dagger Y \Lambda_1^\dagger \Lambda_1\big) \cdot B + \big(\sqrt{X}\sqrt{X}^\dagger Z \Lambda_2^\dagger \Lambda_2\big) \cdot C\,,$$

by the range relations: $\mathrm{Ran}(Y^{\mathrm{T}}) \subset \mathrm{Ran}(\Lambda_1)$, $\mathrm{Ran}(Z^{\mathrm{T}}) \subset \mathrm{Ran}(\Lambda_2)$ and $\mathrm{Ran}([Y, Z]) \subset \mathrm{Ran}(X)$. Also, by (3.1), we have $X \cdot \big(A + \frac{1}{2} B \Lambda_1^2 B^{\mathrm{T}} + \frac{1}{2} C \Lambda_2^2 C^{\mathrm{T}}\big) \le 0$. Hence, by (3.5) and (3.6), the maximisers to (3.4) are given by the set

$$\Big\{(A, B, C) \in \mathcal{O}_\Lambda\,;\, Y = X B \Lambda_1^2\,,\ Z = X C \Lambda_2^2\,,\ X \cdot \Big(A + \frac{1}{2} B \Lambda_1^2 B^{\mathrm{T}} + \frac{1}{2} C \Lambda_2^2 C^{\mathrm{T}}\Big) = 0\Big\}\,, \tag{3.7}$$

and the corresponding supremum is (3.2).

Finally, to characterise the subgradient of J_Λ, by Lemma 2.4, we have that $(A, B, C) \in \partial J_\Lambda(X, Y, Z)$ if and only if $(A, B, C) \in \mathcal{O}_\Lambda$ and $J_\Lambda(X, Y, Z) = X \cdot A + Y \cdot B + Z \cdot C$ holds. Then, (3.3) readily follows from the above argument. For the last statement, we note that $\partial J_\Lambda(X, Y, Z)$ is a singleton if and only if the equations in (3.3) for (A, B, C) are uniquely solvable, which is equivalent to $\Lambda_1 \in \mathbb{S}^k_{++}$, $\Lambda_2 \in \mathbb{S}^m_{++}$ and $X \in \mathbb{S}^n_{++}$. □

Similarly to the unbalanced WFR distance [27, 56, 64], the variables $(X, Y, Z) \in \mathbb{S}^n \times \mathbb{R}^{n\times k} \times \mathbb{R}^{n\times m}$ in the infinitesimal cost $J_\Lambda(X, Y, Z)$ represent the mass, the momentum for the mass transportation and the source for the mass variation, respectively, in our transport problem (see Remark 3.6 and Definition 3.8). In what follows, we assume $m = n$, since the dimensions of the mass $X \in \mathbb{S}^n$ and the source $Z \in \mathbb{R}^{n\times m}$

need to match. We shall also let $\Lambda_2 \in \mathbb{S}^n_{++}$ to avoid technical issues (see Remark 3.10). Now, for a given triplet of measures $\mu := (\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{M}(\mathcal{X}, \mathbb{S}^n \times \mathbb{R}^{n\times k} \times \mathbb{M}^n)$, we define a positive measure $\mathcal{J}_\Lambda(\mu)$ on $\mathcal{X}$ by

$$\mathcal{J}_\Lambda(\mu)(E) := \int_E J_\Lambda\left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right) \mathrm{d}\lambda\,, \tag{3.8}$$

for a measurable set $E \in \mathscr{B}(\mathcal{X})$, where $\lambda \in \mathcal{M}(\mathcal{X}, \mathbb{R}_+)$ is a reference measure such that $|\mu| \ll \lambda$. Thanks to the positive homogeneity of J_Λ by Proposition 3.1, the definition (3.8) of $\mathcal{J}_\Lambda$ is independent of the choice of λ. To alleviate notations, we adopt the following conventions in the rest of this work.

- We define the space $\mathbb{X} := \mathbb{S}^n \times \mathbb{R}^{n\times k} \times \mathbb{M}^n$ and then write $\mathcal{M}(\mathcal{X}, \mathbb{X}) = \mathcal{M}(\mathcal{X}, \mathbb{S}^n \times \mathbb{R}^{n\times k} \times \mathbb{M}^n) = C(\mathcal{X}, \mathbb{X})^*$, where $C(\mathcal{X}, \mathbb{X}) = C(\mathcal{X}, \mathbb{S}^n \times \mathbb{R}^{n\times k} \times \mathbb{M}^n)$.
- We often write μ for $(\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{M}(\mathcal{X}, \mathbb{X})$ for short, which will be clear from the context.
- We write $\mathcal{J}_\Lambda(\mu)(E)$ as $\mathcal{J}_{\Lambda,E}(\mu)$ for short. Then, $\mathcal{J}_{\Lambda,\mathcal{X}}(\mu)$ denotes the total measure $\mathcal{J}_\Lambda(\mu)(\mathcal{X})$.
- We denote by $(G_\lambda, q_\lambda, R_\lambda)$ the density of $(\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{M}(\mathcal{X}, \mathbb{X})$ with respect to a reference measure $\lambda \in \mathcal{M}(\mathcal{X}, \mathbb{R}_+)$ such that $|(\mathsf{G},\mathsf{q},\mathsf{R})| \ll \lambda$. The subscript λ of $(G_\lambda, q_\lambda, R_\lambda)$ will often be omitted for simplicity.
- The generic positive constant C involved in the estimates below may change from line to line.

Definition 3.2. *We define the* Λ*-weighted action functional for a measure* $\mu \in \mathcal{M}(\mathcal{X}, \mathbb{X})$ *by* $\mathcal{J}_{\Lambda,\mathcal{X}}(\mu)$.

By Proposition 3.1 and the formula (3.8), we have the following useful lemma.

Lemma 3.3. *For* $\mu = (\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{M}(\mathcal{X}, \mathbb{X})$ *with* $\mathcal{J}_{\Lambda,\mathcal{X}}(\mu) < +\infty$, *we have* $\mathsf{G} \in \mathcal{M}(\mathcal{X}, \mathbb{S}^n_+)$ *and* $|(\mathsf{q},\mathsf{R})| \ll Tr\mathsf{G}$ *with*

$$G_\lambda \in \mathbb{S}^n_+,\ \operatorname{Ran}([q_\lambda, R_\lambda]) \subset \operatorname{Ran}(G_\lambda),\ \operatorname{Ran}(q_\lambda^{\mathrm{T}}) \subset \operatorname{Ran}(\Lambda_1),\ \operatorname{Ran}(R_\lambda^{\mathrm{T}}) \subset \operatorname{Ran}(\Lambda_2), \quad \lambda\text{-a.e.}\,. \tag{3.9}$$

Proof. By $\mathcal{J}_{\Lambda,\mathcal{X}}(\mu) = \int_{\mathcal{X}} J_\Lambda(\mu_\lambda)\,\mathrm{d}\lambda < +\infty$, $J_\Lambda(\mu_\lambda)$ is finite for λ-a.e. $x \in \mathcal{X}$, where $\mu_\lambda = (G_\lambda, q_\lambda, R_\lambda)$. It means that $\mu_\lambda(x) \in dom(J_\Lambda)$ holds λ-a.e., which immediately gives (3.9) by Proposition 3.1. We next show the absolute continuity of $|\mathsf{q}|$ and $|\mathsf{R}|$ with respect to $Tr\mathsf{G}$, that is, for $E \in \mathscr{B}(\mathcal{X})$ with $Tr\mathsf{G}(E) = 0$, we have $|\mathsf{q}|(E) = |\mathsf{R}|(E) = 0$. For this, we consider two measurable subsets E_1 and E_2 of E with $E = E_1 \cup E_2$:

$$E_1 = \left\{x \in E\,;\, G_\lambda(x) \in \mathbb{S}^n_+ \backslash \{0\}\right\}, \quad E_2 = \{x \in E\,;\, G_\lambda(x) = 0\}\,.$$

By $Tr\mathsf{G}(E_1) = 0$ and $TrG_\lambda > 0$ on E_1 everywhere, we have $\lambda(E_1) = 0$. Then $|\mathsf{q}|(E_1) = 0$ and $|\mathsf{R}|(E_1) = 0$ follows from $|\mathsf{q}|, |\mathsf{R}| \ll \lambda$. Moreover, by (3.9) and $G_\lambda = 0$ on E_2, we have $q_\lambda(x) = 0$ and $R_\lambda(x) = 0$ for λ-a.e. $x \in E_2$. Then it follows that $|\mathsf{q}|(E_2) = 0$ and $|\mathsf{R}|(E_2) = 0$. The proof is complete. □

3.2. Continuity equation

Another key ingredient for the dynamic OT formulation is a matricial continuity equation; see Definition 3.4 below. Let us fix more notations.

- Let $\Omega \subset \mathbb{R}^d$ be a compact set with a nonempty interior, a smooth boundary $\partial\Omega$ and the exterior unit normal vector $\nu = (\nu_1, \ldots, \nu_d)$. We denote by $Q_a^b := [a, b] \times \Omega \subset \mathbb{R}^{1+d}$ with $b > a > 0$ the associated time-space domain. If $[a, b] = [0, 1]$, we simply write it as Q.
- For a function $\Phi(t, x)$ on Q_a^b, we write $\Phi_t(\,\cdot\,) := \Phi(t, \cdot)$ if we regard it as a family of functions $\{\Phi_t\}_{t\in[a,b]}$ in x.
- We denote by $\pi^t : (t, x) \to t$ the projection. We use the subscript # to denote the pushforward by a map. For instance, for a measure μ on Q_a^b, $\pi^t_\#\mu = \mu \circ (\pi^t)^{-1}$ is the pushforward measure on $[a, b]$.

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

- Let X and Y be two Banach spaces. We denote by $\mathcal{L}(X, Y)$ the space of continuous linear operators from X to Y (simply $\mathcal{L}(X)$ if $X = Y$) and by $C_c^\infty(\mathbb{R}^d, X)$ the X-valued smooth functions with compact support. We also need C^k-smooth functions $C^k(\Omega, X)$, where we assume that the derivatives exist in the interior of Ω and can be continuously extended to the boundary. The norm on $C^k(\Omega, X)$ is defined by $\|\Phi\|_{k,\infty} := \sum_{|\alpha|\le k} \sup_{x\in\Omega} \|D^\alpha \Phi(x)\|$. Other similar notations are interpreted accordingly.
- We recall the indicator function of a set A:

$$\chi_A(x) = \begin{cases} 1, & \text{if } x \in A\,, \\ 0, & \text{if } x \notin A\,. \end{cases} \tag{3.10}$$

- We use $\widehat{\cdot}$ to denote the Fourier transform of a function, or the symbol of a constant coefficient linear differential operator.

Let $\mathsf{D}^* : C_c^\infty(\mathbb{R}^d, \mathbb{S}^n) \to C_c^\infty(\mathbb{R}^d, \mathbb{R}^{n\times k})$ be a general first-order constant coefficient linear differential operator satisfying $\mathsf{D}^*(I) = 0$. That is, for a matrix-valued function $\Phi \in C_c^\infty(\mathbb{R}^d, \mathbb{S}^n)$ with components $\{\Phi_{ij}\}_{i,j=1}^n$, we have

$$\mathsf{D}^*(\Phi_{ij}(e_{ij} + e_{ji})) = A_0^{ij}\Phi_{ij}(x) + \sum_{l=1}^d A_l^{ij}\partial_{x_l}\Phi_{ij}(x)\,, \quad i \le j\,, \tag{3.11}$$

for some matrices $\{A_l^{ij}\}_{l=0}^d \subset \mathbb{R}^{n\times k}$, and there holds $\sum_{i=1}^n A_0^{ii} = 0$. Here e_{ij} is the $n \times n$ matrix unit with 1 at the (i, j)-entry. By Fourier transform, the operator D^* can be equivalently characterised by

$$\mathsf{D}^*(\Phi)(x) = \int_{\mathbb{R}^d} \widehat{\mathsf{D}^*}(\xi)\big[\widehat{\Phi}(\xi)\big]e^{\mathrm{i}\xi\cdot x}\,\mathrm{d}\xi\,, \quad \Phi \in C_c^\infty(\mathbb{R}^d, \mathbb{S}^n)\,, \tag{3.12}$$

where $\widehat{\Phi}(\xi)$ is the Fourier transform of Φ:

$$\widehat{\Phi}(\xi) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \Phi(x)e^{-\mathrm{i}\xi\cdot x}\,\mathrm{d}x\,,$$

and $\widehat{\mathsf{D}^*}(\xi) : \mathbb{R}^d \to \mathcal{L}(\mathbb{S}^n, \mathbb{R}^{n\times k})$ is the symbol of D^* such that for any $X \in \mathbb{S}^n$ and $Y \in \mathbb{R}^{n\times k}$, $Y \cdot \widehat{\mathsf{D}^*}(\xi)[X]$ is a first-order polynomial in ξ. We write $\widehat{\mathsf{D}^*}(\xi)$ as the sum of its homogeneous components: $\widehat{\mathsf{D}^*}(\xi) = \widehat{\mathsf{D}_0^*} + \widehat{\mathsf{D}_1^*}(\xi)$, where $\widehat{\mathsf{D}_0^*}$ and $\widehat{\mathsf{D}_1^*}(\xi)$ are homogeneous of degree 0 and 1, respectively: for $X = (X_{ij}) \in \mathbb{S}^n$,

$$\widehat{\mathsf{D}_0^*}[X] = \frac{1}{2}\sum_{i=1}^n A_0^{ii}X_{ii} + \sum_{i<j} A_0^{ij}X_{ij}\,,$$

and

$$\widehat{\mathsf{D}_1^*}(\xi)[X] = \frac{\mathrm{i}}{2}\sum_{l=1}^d\sum_{i=1}^n A_l^{ii}\xi_l X_{ii} + \mathrm{i}\sum_{l=1}^d\sum_{i<j} A_l^{ij}\xi_l X_{ij}\,,$$

with matrices A_l^{ij} given in (3.11). Then, noting that the Fourier transform of I is $\delta_0 I$, it is easy to see that the condition $\mathsf{D}^*(I) = 0$ is equivalent to $\widehat{\mathsf{D}^*}(0)(I) = \widehat{\mathsf{D}_0^*}(I) = \frac{1}{2}\sum_{i=1}^n A_0^{ii} = 0$.

By abuse of notation, we define $\mathsf{D}^*\Phi$ for functions $\Phi(t, x)$ on $\mathbb{R}^{1+d}$ by acting D^* on the spatial variable x. Moreover, we define the operator D as the adjoint operator of $-\mathsf{D}^*$ in the sense of distribution, which can be viewed as a bdivergence operator that maps the momentum to the mass (see equation (3.14)). We similarly denote by D_0 and D_1 the homogeneous parts of degree 0 and 1 of the operator D, respectively.

Example 3.1. *A simple example of* D *is the entry-wise transport, in which case the mass transportation between components is forbidden. To be precise, for* $\mathsf{q} \in \mathcal{M}(Q, \mathbb{R}^{n\times n\times d})$*, we regard* q *as a collection of* $\mathbb{R}^d$*-valued measures* $\{\mathsf{q}_{ij}\}_{i,j=1}^n \subset \mathcal{M}(Q, \mathbb{R}^d)$*, and define*

$$\mathsf{D}(\mathsf{q}) = (\operatorname{div}\mathsf{q})^{\mathrm{sym}} = \frac{\operatorname{div}\mathsf{q} + (\operatorname{div}\mathsf{q})^{\mathrm{T}}}{2}\,,$$

where the standard divergence is applied to each q_{ij}*, i.e.,* $(\mathrm{div}\mathsf{q})_{ij} := \mathrm{div} q_{ij}$*. Then, the adjoint* D^* *is simply given by the gradient that acts on* $\Phi \in C_c^\infty(\mathbb{R}^d, \mathbb{S}^n)$ *component-wisely:* $\mathsf{D}^*\Phi = (\nabla \Phi_{ij})_{ij}$*. More examples with discussion can be found in Section 7.*

Definition 3.4. *A measure* $\mathsf{G} \in \mathcal{M}(Q_a^b, \mathbb{S}^n)$ *connects* $\mathsf{G}_a, \mathsf{G}_b \in \mathcal{M}(\Omega, \mathbb{S}_+^n)$ *over the time interval* $[a, b]$*, if there exists* $(\mathsf{q}, \mathsf{R}) \in \mathcal{M}(Q_a^b, \mathbb{R}^{n\times k} \times \mathbb{M}^n)$ *satisfying the following general matrix-valued continuity equation:*

$$\int_{Q_a^b} \partial_t \Phi \cdot \mathrm{d}\mathsf{G} + \mathsf{D}^*\Phi \cdot \mathrm{d}\mathsf{q} + \Phi \cdot \mathrm{d}\mathsf{R} = \int_\Omega \Phi_b \cdot \mathrm{d}\mathsf{G}_b - \int_\Omega \Phi_a \cdot \mathrm{d}\mathsf{G}_a\,, \quad \forall \Phi \in C^1(Q_a^b, \mathbb{S}^n)\,. \tag{3.13}$$

The measures G_a *and* G_b *are referred to as the initial and final distributions of* G*, respectively. Moreover, we denote by* $\mathcal{CE}([a, b]; \mathsf{G}_a, \mathsf{G}_b)$ *the set of the measures* $(\mathsf{G}, \mathsf{q}, \mathsf{R}) \in \mathcal{M}(Q_a^b, \mathbb{X})$ *satisfying (3.13).*

Remark 3.5. It is easy to derive the distributional equation of (3.13):

$$\partial_t \mathsf{G} + \mathsf{D}\mathsf{q} = \mathsf{R}^{\mathrm{sym}}\,, \tag{3.14}$$

with the measure q satisfying a homogeneous boundary condition on $\partial\Omega$. Indeed, assume that q admits a smooth density q with respect to the Lebesgue measure. Note that for $\mathsf{D}^* = a + \partial_{x_i}$ with $\mathsf{D} = -a + \partial_{x_i}$ ($a \in \mathbb{R}$), a direct integration by parts gives, for smooth real functions f, g on Ω,

$$\int_\Omega ((a + \partial_{x_i})f(x))g(x) + f(x)(-a + \partial_{x_i})g(x)\, \mathrm{d}x = \int_{\partial\Omega} \nu_i f(x) g(x)\, \mathrm{d}x\,.$$

We then have, by linearity and noting $\widehat{\partial_{x_k}} = \mathrm{i}\xi_k$, for a general D^*,

$$\int_\Omega \mathsf{D}q \cdot \Phi + q \cdot \mathsf{D}^*\Phi\, \mathrm{d}x = \int_{\partial\Omega} q \cdot \widehat{\mathsf{D}_1^*}(-\mathrm{i}\nu)(\Phi)\, \mathrm{d}x = \int_{\partial\Omega} \widehat{\mathsf{D}_1}(-\mathrm{i}\nu)(q) \cdot \Phi\, \mathrm{d}x\,, \quad \forall \Phi \in C^1(\Omega, \mathbb{S}^n)\,.$$

It follows that the boundary condition $\widehat{\mathsf{D}_1}(-\mathrm{i}\nu)(q) = 0$ holds for q satisfying (3.13). In the case of $\mathsf{D} = \mathrm{div}$ for $\mathsf{q} \in \mathcal{M}(Q, \mathbb{R}^d)$, we see that $\widehat{\mathsf{D}_1}(-\mathrm{i}\nu)(q) = 0$ is the familiar no-flux boundary condition $\nu \cdot q = 0$.

Remark 3.6. We give an intuitive interpretation of (3.14) as a continuity equation. Recall the homogeneous parts D_0 and D_1 of D with $\mathsf{D}_0 \in \mathcal{L}(\mathbb{R}^{n\times k}, \mathbb{S}^n)$ and D_1 vanishing when acting on constant functions. It allows us to split $\mathsf{D}\mathsf{q}$ into two parts: $\mathsf{D}_0\mathsf{q}$ and $\mathsf{D}_1\mathsf{q}$, where $\mathsf{D}_0\mathsf{q}$ and $\mathsf{D}_1\mathsf{q}$ describe the mass transportation between components of G and the transportation in space, respectively. Moreover, the condition $\mathsf{D}^*(I) = 0$ can be regarded as a conservativity condition in the sense that if $\mathsf{R} = 0$, then $Tr\mathsf{G}_t(\Omega) = Tr\mathsf{G}_0(\Omega)$ for any t; see Proposition 3.13.

The following elementary lemma gives the absolute continuity of the time marginal of G.

Lemma 3.7. *Let* $(\mathsf{G}, \mathsf{q}, \mathsf{R}) \in \mathcal{CE}([a, b]; \mathsf{G}_a, \mathsf{G}_b)$ *with* $\mathsf{G}_a, \mathsf{G}_b \in \mathcal{M}(\Omega, \mathbb{S}_+^n)$*. It holds that* $\pi_\#^t\mathsf{G} \in \mathcal{M}([a, b], \mathbb{S}^n)$ *has the distributional derivative* $(\pi_\#^t\mathsf{R})^{\mathrm{sym}} \in \mathcal{M}([a, b], \mathbb{S}^n)$ *in t. If, further,* $\mathsf{G} \in \mathcal{M}(Q_a^b, \mathbb{S}_+^n)$ *is a positive semi-definite matrix-valued measure over* Q_a^b*, then* $\pi_\#^t|\mathsf{G}| \ll \mathrm{d}t$*.*

Proof. It suffices to consider $[a, b] = [0, 1]$. By (3.13) with test functions $\Phi(t, x) = \phi(t) \in C_c^1((0, 1), \mathbb{S}^n)$, we have

$$\int_0^1 \partial_t \phi \cdot \mathrm{d}\pi_\#^t\mathsf{G} + \phi \cdot \mathrm{d}\pi_\#^t\mathsf{R} = 0\,, \tag{3.15}$$

which implies that $(\pi_\#^t\mathsf{R})^{\mathrm{sym}}$ is the distributional derivative of $\pi_\#^t\mathsf{G}$. Note that $\pi_\#^t\mathsf{G}$ and $\pi_\#^t\mathsf{R}$ are Radon measures (since every finite Borel measure on $[0, 1]$ is regular). There exists a matrix-valued bounded variation function $M(t)$ that generates the Radon measure $\pi_\#^t\mathsf{R}$ [42, Theorem 3.29]. It follows from (3.15) that

$$\mathrm{d}\pi_\#^t\mathsf{G} = (M(t)^{\mathrm{sym}} + C)\, \mathrm{d}t\,, \tag{3.16}$$

for some $C \in \mathbb{S}^n$ [42, Theorem 3.36]. bIf $\mathsf{G} \in \mathcal{M}(Q_a^b, \mathbb{S}_+^n)$, then (3.16) and (2.2) readily give $Tr\pi_\#^t\mathsf{G} \sim |\pi_\#^t\mathsf{G}| \ll \mathrm{d}t$, which further yields $\pi_\#^t|\mathsf{G}| \ll \mathrm{d}t$ by noting $Tr\pi_\#^t\mathsf{G} = \pi_\#^t Tr\mathsf{G} \sim \pi_\#^t|\mathsf{G}|$. □

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

3.3. Weighted Wasserstein–Bures distance

We are now ready to define a class of distances on $\mathcal{M}(\Omega, \mathbb{S}^n_+)$ by minimising the action functional $\mathcal{J}_{\Lambda,\mathcal{Q}}(\mu)$ over the solutions to the continuity equation (3.13).

Definition 3.8. *The weighted Wasserstein–Bures distance between* $\mathsf{G}_0, \mathsf{G}_1 \in \mathcal{M}(\Omega, \mathbb{S}^n_+)$ *is defined by*

$$\mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1) = \inf_{\mu \in \mathcal{CE}([0,1];\, \mathsf{G}_0, \mathsf{G}_1)} \mathcal{J}_{\Lambda,\mathcal{Q}}(\mu). \tag{$\mathcal{P}$}$$

We remark that the quantity $\mathcal{J}_{\Lambda,\mathcal{Q}}(\mu)$ can be understood as the energy of the measure $\mu \in \mathcal{CE}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$. The following a priori estimate shows that $\mathcal{CE}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$ is nonempty and $WB_\Lambda(\mathsf{G}_0, \mathsf{G}_1)$ is always finite, which means that the problem ($\mathcal{P}$) is well defined.

Lemma 3.9. *Given* $\mathsf{G}_0, \mathsf{G}_1 \in \mathcal{M}(\Omega, \mathbb{S}^n_+)$, *let* $\lambda \in \mathcal{M}(\Omega, \mathbb{R}_+)$ *be a reference measure such that* $|\mathsf{G}_0|, |\mathsf{G}_1| \ll \lambda$. *Then there exists* $\mu = (\mathsf{G},0,\mathsf{R}) \in \mathcal{CE}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$ *with finite* $\mathcal{J}_{\Lambda,\mathcal{Q}}(\mu)$. *Moreover, it holds that*

$$\mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1) \le \mathrm{WB}^2_{(0,\Lambda_2)}(\mathsf{G}_0, \mathsf{G}_1) \le 2\|\Lambda_2^{-1}\|^2_{\mathrm{F}} \int_\Omega \left\|\sqrt{G_{1,\lambda}} - \sqrt{G_{0,\lambda}}\right\|^2_{\mathrm{F}} \,\mathrm{d}\lambda\,, \tag{3.17}$$

where $G_{0,\lambda}$ *and* $G_{1,\lambda}$ *are densities of* G_0 *and* G_1 *with respect to* λ.

Proof. We omit the subscript λ of $G_{0,\lambda}$ and $G_{1,\lambda}$ for simplicity. We define measures

$$\mathsf{G} := \left(\sqrt{G_0} + t\left(\sqrt{G_1} - \sqrt{G_0}\right)\right)^2 \mathrm{d}t \otimes \lambda \in \mathcal{M}(Q, \mathbb{S}^n_+)\,,$$

and

$$\mathsf{R} := 2\left(\sqrt{G_0} + t\left(\sqrt{G_1} - \sqrt{G_0}\right)\right)\left(\sqrt{G_1} - \sqrt{G_0}\right) \mathrm{d}t \otimes \lambda \subset \mathcal{M}(Q, \mathbb{M}^n)\,,$$

which satisfies $\mu = (\mathsf{G},0,\mathsf{R}) \in \mathcal{CE}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$ and $\mathrm{Ran}\left(\frac{\mathrm{d}\mathsf{R}}{\mathrm{d}t\otimes\lambda}\right) \subset \mathrm{Ran}\left(\frac{\mathrm{d}\mathsf{G}}{\mathrm{d}t\otimes\lambda}\right)$ for $\mathrm{d}t \otimes \lambda$-a.e. Moreover, we note

$$\mathrm{Ran}\left(\sqrt{G_1} - \sqrt{G_0}\right) \subset \mathrm{Ran}\left(\sqrt{G_0} + t\left(\sqrt{G_1} - \sqrt{G_0}\right)\right)\,, \quad t \in (0, 1)\,,$$

from the relation: $\mathrm{Ker}\left(\sqrt{G_0} + t(\sqrt{G_1} - \sqrt{G_0})\right) = \mathrm{Ker}\left(\sqrt{G_0}\right) \cap \mathrm{Ker}\left(\sqrt{G_1}\right) \subset \mathrm{Ker}\left(\sqrt{G_1} - \sqrt{G_0}\right)$. Then, we compute

$$\mathcal{J}_{\Lambda,\mathcal{Q}}(\mu) = 2\int_\Omega \left\|\left(\sqrt{G_1} - \sqrt{G_0}\right)\Lambda_2^{-1}\right\|^2_{\mathrm{F}} \mathrm{d}\lambda\,, \tag{3.18}$$

for μ defined above. The proof is completed by the submultiplicativity of the Frobenius norm. □

Remark 3.10. The proof of Lemma 3.9 uses $\mathrm{Ran}(\Lambda_2) = \mathbb{R}^n$ from the assumption $\Lambda_2 \in \mathbb{S}^n_{++}$ we made before (3.8). If we only assume $\Lambda_2 \in \mathbb{S}^n_+$, the distance WB_Λ may be only well-defined (i.e., finite) on a subset of $\mathcal{M}(\Omega, \mathbb{S}^n_+)$.

Remark 3.11. $\mathrm{WB}_{(0,\Lambda_2)}$ *is the matricial Hellinger distance* d_H *in [73, Definition 4.1], up to a transformation. Indeed, recalling Lemma 3.3, we have that if* $\Lambda_1 = 0$, *then* q *must be zero and* ($\mathcal{P}$) *reduces to*

$$\mathrm{WB}^2_{(0,\Lambda_2)}(\mathsf{G}_0, \mathsf{G}_1) = \inf\{\mathcal{J}_{(0,\Lambda_2),\mathcal{Q}}(\mu)\,;\, \mu = (\mathsf{G},0,\mathsf{R}) \in \mathcal{CE}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)\}. \tag{3.19}$$

For a given $S \in \mathbb{S}^n_{++}$, *we introduce a linear map* $g_S(A) := SAS\colon \mathbb{S}^n_+ \to \mathbb{S}^n_+$ *with the inverse* $g_{S^{-1}}$. *It is easy to see that* $(\mathsf{G},0,\mathsf{R}) \in \mathcal{CE}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$ *if and only if* $(g_{\Lambda_2^{-1}}(\mathsf{G}), 0, g_{\Lambda_2^{-1}}(\mathsf{R})) \in \mathcal{CE}([0, 1]; g_{\Lambda_2^{-1}}(\mathsf{G}_0), g_{\Lambda_2^{-1}}(\mathsf{G}_1))$, *and there holds* $\mathcal{J}_{(0,\Lambda_2),\mathcal{Q}}((\mathsf{G},0,\mathsf{R})) = \mathcal{J}_{(0,I),\mathcal{Q}}(g_{\Lambda_2^{-1}}(\mathsf{G}), 0, g_{\Lambda_2^{-1}}(\mathsf{R}))$. *Therefore, we have*

$$\mathrm{WB}_{(0,\Lambda_2)}(\mathsf{G}_0, \mathsf{G}_1) = \mathrm{WB}_{(0,I)}(g_{\Lambda_2^{-1}}(\mathsf{G}_0), g_{\Lambda_2^{-1}}(\mathsf{G}_1))\,.$$

From [73, Definition 4.1] and Theorem 4.5 below, one can see that $\mathrm{WB}_{(0,I)}$ *is nothing else than the convex formulation of the Hellinger distance* d_H, *up to a constant. We refer the readers to [73, Lemma 4.3 and*

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

Theorem 2] for the properties of the Hellinger distance and its relation with the Bures-Wasserstein distance on $\mathbb{S}_+^n$ [10].

3.4. A priori estimate

Thanks to Lemma 3.9, the optimisation ($\mathcal{P}$) can be equivalently taken over the following set:

$$\mathcal{CE}_\infty([0,1]; \mathsf{G}_0, \mathsf{G}_1) := \mathcal{CE}([0,1]; \mathsf{G}_0, \mathsf{G}_1) \bigcap \{\mu \in \mathcal{M}(Q, \mathbb{X}); \mathcal{J}_{\Lambda,Q}(\mu) < +\infty\}\,.$$

Before we proceed, we give some auxiliary results. First, we introduce

$$\mathcal{J}^*_{\Lambda,\mathcal{X}}(\mathsf{G}, u, W) := \frac{1}{2}\|(u\Lambda_1, W\Lambda_2)\|^2_{L^2_{\mathsf{G}}(\mathcal{X})} \quad \text{on } \mathcal{M}(\mathcal{X}, \mathbb{S}^n_+) \times C(\mathcal{X}, \mathbb{R}^{n\times k} \times \mathbb{M}^n)\,, \tag{3.20}$$

where $\|\cdot\|_{L^2_{\mathsf{G}}(\mathcal{X})}$ is defined by (2.3). By an argument similar to the one for Lemma 4.1 below, we have that the conjugate function (2.6) of $\mathcal{J}^*_{\Lambda,\mathcal{X}}(\mathsf{G}, u, W)$ with respect to (u, W) is exactly $\mathcal{J}_{\Lambda,\mathcal{X}}(\mathsf{G}, \mathsf{q}, \mathsf{R})$. Moreover, there holds

$$\mathcal{J}_{\Lambda,\mathcal{X}}(\mathsf{G}, \mathsf{q}, \mathsf{R}) = \sup_{(u,W)\in L^\infty_{|(\mathsf{G},\mathsf{q},\mathsf{R})|}(\mathcal{X},\mathbb{R}^{n\times k}\times\mathbb{M}^n)} \langle(\mathsf{q}, \mathsf{R}), (u, W)\rangle_{\mathcal{X}} - \mathcal{J}^*_{\Lambda,\mathcal{X}}(\mathsf{G}, u, W)\,. \tag{3.21}$$

Since $\mathcal{J}_{\Lambda,\mathcal{X}}(\mathsf{G}, \mathsf{q}, \mathsf{R})$ and $\mathcal{J}^*_{\Lambda,\mathcal{X}}(\mathsf{G}, u, W)$ are homogeneous of degree 2 in (q, R) and (u, W), respectively, by (3.21), it holds that for $(\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{M}(\mathcal{X}, \mathbb{X})$ and $(u, W) \in L^\infty_{|(\mathsf{G},\mathsf{q},\mathsf{R})|}(\mathcal{X}, \mathbb{R}^{n\times k} \times \mathbb{M}^n)$,

$$\langle(\mathsf{q}, \mathsf{R}), (u, W)\rangle_{\mathcal{X}} \le \gamma^{-2}\mathcal{J}_{\Lambda,\mathcal{X}}(\mathsf{G}, \mathsf{q}, \mathsf{R}) + \gamma^2\mathcal{J}^*_{\Lambda,\mathcal{X}}(\mathsf{G}, u, W)\,, \quad \forall\gamma > 0\,. \tag{3.22}$$

We minimise the right-hand side of (3.22) with respect to γ and obtain

$$\langle(\mathsf{q}, \mathsf{R}), (u, W)\rangle_{\mathcal{X}} \le 2\sqrt{\mathcal{J}_{\Lambda,\mathcal{X}}(\mathsf{G}, \mathsf{q}, \mathsf{R})\mathcal{J}^*_{\Lambda,\mathcal{X}}(\mathsf{G}, u, W)}\,, \tag{3.23}$$

where we have used non-negativity of $\mathcal{J}_{\Lambda,\mathcal{X}}$ and $\mathcal{J}^*_{\Lambda,\mathcal{X}}$.

Second, we observe from formulas (3.2) and (3.8) and Lemmas 2.3 and 3.3 that for $\mu = (\mathsf{G}, \mathsf{q}, \mathsf{R}) \in \mathcal{M}(\mathcal{X}, \mathbb{X})$ with $\mathcal{J}_{\Lambda,\mathcal{X}}(\mu) < +\infty$, the functions $G^\dagger_\lambda q_\lambda \Lambda_1^\dagger$ and $G^\dagger_\lambda R_\lambda \Lambda_2^{-1}$ are well defined, Borel measurable and independent of the reference measure λ (hence we omit the subscript λ in the sequel for simplicity), and there holds

$$\mathcal{J}_{\Lambda,\mathcal{X}}(\mu) = \frac{1}{2}\|G^\dagger q\Lambda_1^\dagger\|^2_{L^2_{\mathsf{G}}(\mathcal{X})} + \frac{1}{2}\|G^\dagger R\Lambda_2^{-1}\|^2_{L^2_{\mathsf{G}}(\mathcal{X})} < +\infty\,. \tag{3.24}$$

We now give useful a priori bounds for measures q and R.

Lemma 3.12. *For $\mu = (\mathsf{G}, \mathsf{q}, \mathsf{R}) \in \mathcal{M}(\mathcal{X}, \mathbb{X})$ with $\mathcal{J}_{\Lambda,\mathcal{X}}(\mu) < +\infty$, it holds that for $E \in \mathscr{B}(\mathcal{X})$,*

$$|\mathsf{q}|(E) \le \sqrt{Tr\mathsf{G}(E)}\,\|\Lambda_1\|_{\mathrm{F}}\|G^\dagger q\Lambda_1^\dagger\|_{L^2_{\mathsf{G}}(E)}\,, \quad |\mathsf{R}|(E) \le \sqrt{Tr\mathsf{G}(E)}\,\|\Lambda_2\|_{\mathrm{F}}\|G^\dagger R\Lambda_2^{-1}\|_{L^2_{\mathsf{G}}(E)}\,. \tag{3.25}$$

Proof. Recall that there exist bounded measurable functions σ_q and σ_R with $\|\sigma_q\|_{\mathrm{F}} = \|\sigma_R\|_{\mathrm{F}} = 1$ such that $\mathrm{d}\mathsf{q} = \sigma_q\,\mathrm{d}|\mathsf{q}|$ and $\mathrm{d}\mathsf{R} = \sigma_R\,\mathrm{d}|\mathsf{R}|$. Taking $\mathsf{R} = 0$ and $(u, W) = (\chi_E\sigma_q, 0)$ in (3.23) for $E \in \mathscr{B}(\mathcal{X})$, we obtain

$$|\mathsf{q}|(E) = \int_E u\cdot \mathrm{d}\mathsf{q} \le 2\sqrt{\mathcal{J}_{\Lambda,E}(\mathsf{G}, \mathsf{q}, 0)\mathcal{J}^*_{\Lambda,E}(\mathsf{G}, u, 0)} \le \sqrt{Tr\mathsf{G}(E)\|\Lambda_1\|^2_{\mathrm{F}}}\|G^\dagger q\Lambda_1^\dagger\|_{L^2_{\mathsf{G}}(E)}\,,$$

by (3.24) and the following estimate derived from (3.20) and (2.4):

$$\mathcal{J}^*_{\Lambda,E}(\mathsf{G}, u, W) \le \frac{1}{2}Tr\mathsf{G}(E)\|\Lambda_1\|^2_{\mathrm{F}}\,.$$

Similarly, by taking $\mathsf{q} = 0$ and $(u, W) = (0, \chi_E\sigma_R)$ in (3.23), we obtain the estimate for R in (3.25). □

With the help of the above lemma, the following proposition holds.

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

Proposition 3.13. *Let* $\mu = (\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{CE}_\infty([0,1];\mathsf{G}_0,\mathsf{G}_1)$ *with* $\mathsf{G}_0, \mathsf{G}_1 \in \mathcal{M}(\Omega, \mathbb{S}^n_+)$. *Then,*

(i) $\mathsf{G} \in \mathcal{M}(Q, \mathbb{S}^n_+)$ *and* $\pi^t_\# |\mathsf{G}| \ll \mathrm{d}t$. *Moreover,* μ *can be disintegrated as*

$$\mu = \int_0^1 \delta_t \otimes (\mathsf{G}_t, \mathsf{q}_t, \mathsf{R}_t)\, \mathrm{d}t\,, \tag{3.26}$$

where $(\mathsf{G}_t, \mathsf{q}_t, \mathsf{R}_t) \in \mathcal{M}(\Omega, \mathbb{X})$ *for* $\mathrm{d}t$*-a.e.* $t \in [0,1]$.

(ii) There exists a weak continuous curve* $\{\widetilde{\mathsf{G}}\}_{t\in[0,1]}$ *in* $\mathcal{M}(\Omega, \mathbb{S}^n_+)$ *such that* $\mathsf{G}_t = \widetilde{\mathsf{G}}_t$ *for a.e.* $t \in [0,1]$ *and, for any interval* $[t_0, t_1] \subset [0,1]$, *it holds that*

$$\int_{Q_{t_0}^{t_1}} \partial_t \Phi \cdot \mathrm{d}\mathsf{G} + \mathsf{D}^*\Phi \cdot \mathrm{d}\mathsf{q} + \Phi \cdot \mathrm{d}\mathsf{R} = \int_\Omega \Phi_{t_1} \cdot \mathrm{d}\widetilde{\mathsf{G}}_{t_1} - \int_\Omega \Phi_{t_0} \cdot \mathrm{d}\widetilde{\mathsf{G}}_{t_0}\,, \quad \forall \Phi \in C^1(Q_{t_0}^{t_1}, \mathbb{S}^n)\,. \tag{3.27}$$

Moreover, there holds, for some $C > 0$,

$$Tr\widetilde{\mathsf{G}}_t(\Omega) \le C\left(Tr\mathsf{G}_0(\Omega) + \|G^\dagger R \Lambda_2^{-1}\|^2_{L^2_{\mathsf{G}}(Q)} \|\Lambda_2\|_{\mathrm{F}}^2\right), \quad \forall t \in [0,1]\,. \tag{3.28}$$

Remark 3.14. By Proposition 3.13, we can identify a measure $\mu = (\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{CE}_\infty([0,1];\mathsf{G}_0,\mathsf{G}_1)$ with a family of measures $\{\mu_t = (\mathsf{G}_t, \mathsf{q}_t, \mathsf{R}_t)\}_{t\in[0,1]}$ in $\mathcal{M}(\Omega, \mathbb{X})$ via the disintegration (3.26), where G_t is weak* continuous. We also remark that one can alternatively define the matrix-valued continuity equation (3.13) by testing against functions $\Phi \in C^1(Q, \mathbb{S}^n)$ compactly supported in $(0,1) \times \Omega$ as in [1, Chapter 8] (in this case the right-hand side of (3.13) vanishes), and consider its solution $\mu = (\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{M}(Q, \mathbb{X})$ with finite energy $\mathcal{J}_{\Lambda,Q}(\mu) < +\infty$. In this setting, a similar analysis by disintegration shows that G still has the weak* continuous representation $\{\mathsf{G}_t\}_{t\in[0,1]}$, and then the initial and final distributions G_0 and G_1 can be obtained from the limits as $t \to 0$ and $t \to 1$ of G_t, respectively. In this work, we always stick to Definition 3.4 with temporal boundary conditions G_0 and G_1 to avoid any confusion.

Proof. (i) First, note from [1, Theorem 5.3.1] that μ can be disintegrated with respect to $\nu = \pi^t_\#|\mu|$ as $\mu = \int_0^1 \delta_t \otimes \mu_t \,\mathrm{d}\nu$, where $\mu_t \in \mathcal{M}(\Omega, \mathbb{X})$ for ν-a.e. $t \in [0,1]$. Then, by Lemmas 3.3 and 3.7, we have $\mathsf{G} \in \mathcal{M}(Q, \mathbb{S}^n_+)$ and $\nu \ll \pi^t_\#|\mathsf{G}| \ll \mathrm{d}t$ on $[0,1]$, which allows us to define $\widetilde{\mu}_t := \mu_t \frac{\mathrm{d}\nu}{\mathrm{d}t}$ and disintegrate μ as $\mu = \int_0^1 \delta_t \otimes \widetilde{\mu}_t \,\mathrm{d}t$.

(ii) Consider test functions $\Phi = a(t)\Psi(x)$ in (3.13) with $a(t) \in C^1_c((0,1), \mathbb{R})$ and $\Psi(x) \in C^1(\Omega, \mathbb{S}^n)$. Then, by (3.26), $\int_\Omega \Psi \cdot \mathrm{d}\mathsf{G}_t$ is absolutely continuous in t with the weak derivative:

$$\partial_t \langle \mathsf{G}_t, \Psi \rangle_\Omega = \langle \mathsf{q}_t, \mathsf{D}^*\Psi \rangle_\Omega + \langle \mathsf{R}_t, \Psi \rangle_\Omega\,. \tag{3.29}$$

Letting $\Psi = I$ in (3.29), we obtain $\partial_t Tr\mathsf{G}_t(\Omega) = Tr\mathsf{R}^{\mathrm{sym}}_t(\Omega)$ a.e. by $\mathsf{D}^*(I) = 0$, which implies that there exists a nonnegative function $m(t) \in C([0,1], \mathbb{R})$ such that $Tr\mathsf{G}_t(\Omega) = m(t)$ a.e. on $[0,1]$ and

$$m(t) - m(s) = \int_s^t Tr\mathsf{R}^{\mathrm{sym}}_\tau(\Omega)\,\mathrm{d}\tau\,, \quad \forall 0 \le s \le t \le 1\,. \tag{3.30}$$

By Lemma 3.12, it follows from (3.30) that, from some $C > 0$,

$$|m(t) - m(s)| \le C|\mathsf{R}|(Q) \le C\sqrt{Tr\mathsf{G}(Q)}\|\Lambda_2\|_{\mathrm{F}} \|G^\dagger R \Lambda_2^{-1}\|_{L^2_{\mathsf{G}}(Q)}\,. \tag{3.31}$$

We choose t_0 such that $m(t_0) = \max_{t\in[0,1]} m(t)$. Then (3.31) implies

$$m(t_0) \le m(0) + C\sqrt{m(t_0)}\|\Lambda_2\|_{\mathrm{F}} \|G^\dagger R \Lambda_2^{-1}\|_{L^2_{\mathsf{G}}(Q)}\,,$$

which further gives, by an elementary calculation,

$$\left(m(t_0)^{1/2} - \frac{C}{2}\|G^\dagger R \Lambda_2^{-1}\|_{L^2_{\mathsf{G}}(Q)} \|\Lambda_2\|_{\mathrm{F}}\right)^2 \le m(0) + \frac{C^2}{4}\|G^\dagger R \Lambda_2^{-1}\|^2_{L^2_{\mathsf{G}}(Q)} \|\Lambda_2\|^2_{\mathrm{F}}\,. \tag{3.32}$$

Then we have

$$m(t) \le C\left(m(0) + \|G^\dagger R \Lambda_2^{-1}\|^2_{L^2_{\mathsf{G}}(Q)} \|\Lambda_2\|^2_{\mathrm{F}}\right). \tag{3.33}$$

With the above estimates, the existence of a weak* continuous representative of G_t and the formula (3.27) can be proved similarly to [1, Lemma 8.1.2]. We sketch the argument for completeness.

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

By (3.25) and (3.33), as well as (3.29), there exists a subset $E \in [0,1]$ of Lebesgue measure zero such that $Tr\mathsf{G}_t(\Omega) = m(t)$ on $[0,1]\backslash E$, and there holds, for any $t, s \in [0,1]\backslash E$ with $s < t$ and $\Psi \in C^1(\Omega, \mathbb{S}^n)$,

$$\begin{aligned} |\langle \mathsf{G}_t, \Psi\rangle_\Omega - \langle \mathsf{G}_s, \Psi\rangle_\Omega| &\le C\|\Psi\|_{1,\infty}\big(|\mathsf{q}|(Q_s^t) + |\mathsf{R}|(Q_s^t)\big) \\ &\le C|t-s|^{1/2}\big(m(0) + \|G^\dagger q \Lambda_1^\dagger\|^2_{L^2_{\mathsf{G}}(Q)}\|\Lambda_1\|^2_{\mathrm{F}} + \|G^\dagger R \Lambda_2^{-1}\|^2_{L^2_{\mathsf{G}}(Q)}\|\Lambda_2\|^2_{\mathrm{F}}\big)\|\Psi\|_{1,\infty}\,. \end{aligned} \tag{3.34}$$

The estimate (3.34) allows us to uniquely extend $\{\mathsf{G}_t\}_{t\in[0,1]\backslash E}$ to a weak* continuous curve $\{\widetilde{\mathsf{G}}_t\}_{t\in[0,1]}$ in $C^1(\Omega,\mathbb{S}^n)^*$. Then, by the density of $C^1(\Omega,\mathbb{S}^n)$ in $C(\Omega,\mathbb{S}^n)$ and the boundedness (3.33) of $\{Tr\widetilde{\mathsf{G}}_t(\Omega)\}_{t\in[0,1]}$, the curve $\{\widetilde{\mathsf{G}}_t\}_{t\in[0,1]}$ is also weak* continuous in $\mathcal{M}(\Omega,\mathbb{S}^n)$. The formula (3.27) follows from taking test functions $\Phi_\varepsilon(x,t) = \eta_\varepsilon(t)\Phi(t,x)$ in (3.13), where $\Phi \in C^1(Q,\mathbb{S}^n)$ and $\eta_\varepsilon \in C_c^\infty((t_0,t_1),\mathbb{R})$ with $0 \le \eta_\varepsilon \le 1$, $\lim_{\varepsilon\to 0}\eta_\varepsilon(t) = \chi_{(t_0,t_1)}(t)$ pointwisely and $\lim_{\varepsilon\to0}\eta'_\varepsilon = \delta_{t_0} - \delta_{t_1}$ in the distributional sense. Recalling $Tr\mathsf{G}_t(\Omega) = m(t)$ a.e., by the weak* continuity of $\widetilde{\mathsf{G}}_t$, we have $Tr\widetilde{\mathsf{G}}_t = m(t)$. Then, the estimate (3.28) follows from (3.33). □

3.5. Time and space scaling

By writing $\mathcal{J}_{\Lambda,Q}(\mu) = \int_0^1 \mathcal{J}_{\Lambda,\Omega}(\mu_t)\,\mathrm{d}t$ for $\mu \in \mathcal{CE}_\infty([0,1];\mathsf{G}_0,\mathsf{G}_1)$, the following Lemma is a simple consequence of the change of variable.

Lemma 3.15. *Let $\mu \in \mathcal{CE}_\infty([0,1];\mathsf{G}_0,\mathsf{G}_1)$. It holds that*

1. *Let $\mathsf{s}(t) : [0,1] \to [a,b]$ be a strictly increasing absolutely continuous map with an absolutely continuous inverse: $\mathsf{t} = \mathsf{s}^{-1}$. Then $\widetilde{\mu} := \int_a^b \delta_s \otimes (\mathsf{G}_{\mathsf{t}(s)}, \mathsf{t}'(s)\mathsf{q}_{\mathsf{t}(s)}, \mathsf{t}'(s)\mathsf{R}_{\mathsf{t}(s)})\,\mathrm{d}s \in \mathcal{CE}([a,b];\mathsf{G}_0,\mathsf{G}_1)$. Moreover, we have*

$$\int_0^1 \mathsf{t}'(\mathsf{s}(t))\mathcal{J}_{\Lambda,\Omega}(\mu_t)\,\mathrm{d}t = \int_a^b \mathcal{J}_{\Lambda,\Omega}(\widetilde{\mu}_s)\,\mathrm{d}s\,. \tag{3.35}$$

2. *Let T be a diffeomorphism on $\mathbb{R}^d$ mapping from Ω to $T(\Omega)$ and suppose that there exists $\mathcal{T}_{\mathsf{D}^*}(x) : \Omega \to \mathcal{L}(\mathbb{R}^{n\times k})$ such that for $\Phi \in C_c^\infty(\mathbb{R}^d,\mathbb{S}^n)$,*

$$\mathcal{T}_{\mathsf{D}^*}[(\mathsf{D}^*\Phi)\circ T] := \mathsf{D}^*(\Phi\circ T)\,. \tag{3.36}$$

Then $\widetilde{\mu} := \int_0^1 \delta_t \otimes T_\#(\mathsf{G}_t, \mathcal{T}_{\mathsf{D}}\mathsf{q}_t, \mathsf{R}_t)\,\mathrm{d}t \in \mathcal{CE}([0,1]; T_\#\mathsf{G}_0, T_\#\mathsf{G}_1)$ on $T(\Omega)$, where $T_\#(\cdot)$ denotes the pushforward measure by T, and $\mathcal{T}_{\mathsf{D}}$ is the transpose of $\mathcal{T}_{\mathsf{D}^*}$ defined via $(\mathcal{T}_{\mathsf{D}}q)\cdot p = q\cdot(\mathcal{T}_{\mathsf{D}^*}p)$, $\forall p, q \in \mathbb{R}^{n\times k}$.

Remark 3.16. The condition (3.36) is nontrivial and necessary for the second statement. Indeed, there holds

$$\mathsf{D}^*(\Phi\circ T) = \int_{\mathbb{R}^d} \widehat{\mathsf{D}^*}(\xi\cdot\nabla T(x))\big[\widehat{\Phi}(\xi)\big]e^{\mathrm{i}\xi\cdot T(x)}\,\mathrm{d}\xi,$$

by Fourier transform, where $(\xi\cdot\nabla T(x))_j = \xi\cdot\partial_j T(x)$. It follows that (3.36) is equivalent to a separation of variables: $\widehat{\mathsf{D}^*}(\xi\cdot\nabla T(x)) = \mathcal{T}_{\mathsf{D}^*}(x)\circ\widehat{\mathsf{D}^*}(\xi)$. A sufficient condition for (3.36) is that $\widehat{\mathsf{D}^*}$ is homogeneous of degree 0, or homogeneous of degree 1 with $T(x) = ax + b$ for $a \neq 0 \in \mathbb{R}$ and $b \in \mathbb{R}^d$, which is enough for our purposes.

Remark 3.17. We connect the weight matrix Λ_1 and the space scaling. Let us consider $\mu \in \mathcal{CE}_\infty([0,1];\mathsf{G}_0,\mathsf{G}_1)$ and D^* be homogeneous of degree one for simplicity. Define $T(x) = ax : \Omega \to a\Omega$ and $\mathcal{T}_{\mathsf{D}} = aI$. By Lemma 3.15, we have $\widetilde{\mu} := \int_0^1 \delta_t \otimes T_\#(\mathsf{G}_t, a\mathsf{q}_t, \mathsf{R}_t)\,\mathrm{d}t \in \mathcal{CE}_\infty([0,1]; T_\#\mathsf{G}_0, T_\#\mathsf{G}_1)$. Then, a direct computation gives

$$\mathcal{J}_{\Lambda,[0,1]\times a\Omega}(\widetilde{\mu}) = \int_0^1 \mathcal{J}_{(a^{-1}\Lambda_1,\Lambda_2),a\Omega}(T_\#(\mathsf{G}_t,\mathsf{q}_t,\mathsf{R}_t))\,\mathrm{d}t = \int_0^1 \mathcal{J}_{(a^{-1}\Lambda_1,\Lambda_2),\Omega}(\mu_t)\,\mathrm{d}t = \mathcal{J}_{(a^{-1}\Lambda_1,\Lambda_2),Q}(\mu).$$

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

Using Lemma 3.15 with $\mathsf{s}(t) = (b-a)t + a : [0,1] \to [a,b]$, $b > a > 0$, we see that for $\mu \in \mathcal{CE}_\infty([0,1]; \mathsf{G}_0, \mathsf{G}_1)$, there exists $\widetilde{\mu} \in \mathcal{CE}_\infty([a,b]; \mathsf{G}_0, \mathsf{G}_1)$ such that

$$\int_0^1 \mathcal{J}_{\Lambda,\Omega}(\mu_t)\,\mathrm{d}t = (b-a)\int_a^b \mathcal{J}_{\Lambda,\Omega}(\widetilde{\mu}_t)\,\mathrm{d}t\,,$$

and vice versa, which gives the equivalent characterisation of WB_Λ:

$$\mathrm{WB}^2_\Lambda(\mathsf{G}_0,\mathsf{G}_1) = \inf_{\mathcal{CE}_\infty([a,b];\mathsf{G}_0,\mathsf{G}_1)} (b-a)\int_a^b \mathcal{J}_{\Lambda,\Omega}(\mu_t)\,\mathrm{d}t\,, \quad \mathsf{G}_0,\mathsf{G}_1 \in \mathcal{M}(\Omega, \mathbb{S}^n_+)\,. \tag{$\mathcal{P}'$}$$

3.6. Compactness

We end the discussion of basic properties of $\mathcal{CE}_\infty([0,1]; \mathsf{G}_0, \mathsf{G}_1)$ with a compactness result.

Proposition 3.18. *Let $\mu^n = (\mathsf{G}^n, \mathsf{q}^n, \mathsf{R}^n) \in \mathcal{CE}_\infty([0,1]; \mathsf{G}^n_0, \mathsf{G}^n_1)$, $n \ge 1$, be a sequence of measures satisfying*

$$m := \sup_{n\in\mathbb{N}} Tr(\mathsf{G}^n_0) < +\infty\,, \quad M := \sup_{n\in\mathbb{N}} \mathcal{J}_{\Lambda,Q}(\mu^n) < +\infty\,. \tag{3.37}$$

Then there exists a subsequence, still denoted by μ^n, and a measure $\mu = (\mathsf{G}, \mathsf{q}, \mathsf{R}) \in \mathcal{CE}_\infty([0,1]; \mathsf{G}_0, \mathsf{G}_1)$ such that for every $t \in [0,1]$, G^n_t weak converges to G_t in $\mathcal{M}(\Omega, \mathbb{S}^n)$, and $(\mathsf{q}^n, \mathsf{R}^n)$ weak* converges to (q,R) in $\mathcal{M}(Q, \mathbb{R}^{n\times k} \times \mathbb{M}^n)$. Moreover, it holds that, for $0 \le a < b \le 1$,*

$$\mathcal{J}_{\Lambda,Q^b_a}(\mu) \le \liminf_{n\to\infty} \mathcal{J}_{\Lambda,Q^b_a}(\mu^n)\,. \tag{3.38}$$

Proof. By (3.37), up to a subsequence, we can let G^n_0 weak* converge to some $\mathsf{G}_0 \in \mathcal{M}(\Omega, \mathbb{S}^n_+)$. It is also clear from a priori estimates (3.25) and (3.28), as well as the assumption (3.37), that $\{\mu^n\}_{n\in\mathbb{N}}$ is bounded in $\mathcal{M}(Q, \mathbb{X})$. Hence, there exists a subsequence of $\{\mu^n\}_{n\in\mathbb{N}}$, still indexed by n, weak* converging to some $\mu \in \mathcal{M}(Q, \mathbb{X})$. We next prove that the restriction of μ^n on Q^b_a, i.e., $\mu^n|_{Q^b_a}$, weak* converges to $\mu|_{Q^b_a}$ in $\mathcal{M}(Q^b_a, \mathbb{X})$ for any $0 \le a \le b \le 1$. For this, again by (3.25) and (3.28), we have, for some $C > 0$,

$$|\mu^n|([t_0, t_1] \times \Omega) \le C|t_1 - t_0|^{1/2}\,, \quad \forall 0 \le t_0 \le t_1 \le 1\,, \tag{3.39}$$

which also holds for μ. Let $\eta(t)$ be a smooth function, compactly supported in $[a, b]$, with $|\eta(t)| \le 1$ and $\eta = 1$ on $[a+\varepsilon, b-\varepsilon]$ for some small ε. Then, for any $\Xi \in C(Q^b_a, \mathbb{X})$, we define $\widetilde{\Xi}(t,x) = \eta(t)\Xi(t,x) \in C(Q, \mathbb{X})$. The following estimate readily follows from the properties of η and the estimate (3.39):

$$\left|\langle \mu^n, \Xi\rangle_{Q^b_a} - \langle \mu, \Xi\rangle_{Q^b_a}\right| \le \left|\langle \mu^n, \widetilde{\Xi}\rangle_Q - \langle \mu, \widetilde{\Xi}\rangle_Q\right| + C\varepsilon^{1/2}\,.$$

Since μ^n weak* converges to μ in $\mathcal{M}(Q, \mathbb{X})$ and ε is arbitrary, we have $\left|\langle \mu^n, \Xi\rangle_{Q^b_a} - \langle \mu, \Xi\rangle_{Q^b_a}\right| \to 0$ as $n \to \infty$ for $\Xi \in C(Q^b_a, \mathbb{X})$. Then, (3.38) follows from the lower semicontinuity of $\mathcal{J}_{\Lambda,Q^b_a}(\mu)$. We now show the weak* convergence of G^n_t for every $t \in [0,1]$. We note, by taking $\Phi(s,x) = \chi_{[0,t]}(s)\Psi(x)$ in (3.27) with $\Psi(x) \in C^1(\Omega, \mathbb{S}^n)$,

$$\int_0^t \left(\int_\Omega \mathrm{D}^*\Psi \cdot \mathrm{d}\mathsf{q}^n_s + \int_\Omega \Psi \cdot \mathrm{d}\mathsf{R}^n_s\right)\mathrm{d}s = \int_\Omega \Psi \cdot \mathrm{d}\mathsf{G}^n_t - \int_\Omega \Psi \cdot \mathrm{d}\mathsf{G}^n_0\,, \quad \forall \Psi \in C^1(\Omega, \mathbb{S}^n)\,.$$

Then, using the weak* convergences of G^n_0 in $\mathcal{M}(\Omega, \mathbb{S}^n)$ and $(\mathsf{q}^n, \mathsf{R}^n)|_{Q^t_0}$ in $\mathcal{M}(Q^t_0, \mathbb{R}^{n\times k} \times \mathbb{M}^n)$, we get the convergence of $\langle \mathsf{G}^n_t, \Psi\rangle_\Omega$ as $n \to \infty$. The proof is completed by the density of $C^1(\Omega, \mathbb{S}^n)$ in $C(\Omega, \mathbb{S}^n)$ and the uniform boundedness of $Tr\mathsf{G}^n_t(\Omega)$ with respect to n from (3.28). □

4. Properties of weighted Wasserstein–Bures metrics

This section is devoted to the investigation of the convex optimisation problem ($\mathcal{P}$). We shall first show the existence of the minimiser and derive the corresponding optimality condition. We then explore its

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

primal-dual formulations in more detail, which will lead to a Riemannian interpretation of WB_Λ in Section 5. Finally, we consider the dependence of WB_Λ on the weight matrix Λ.

4.1. Existence of minimiser and optimality condition

For our purpose, let us first define the Lagrangian of ($\mathcal{P}$) with the multiplier $\Phi \in C^1(Q, \mathbb{S}^n)$:

$$\mathcal{L}(\mu, \Phi) := \mathcal{J}_{\Lambda,Q}(\mu) - \langle \mu, (\partial_t \Phi, \mathsf{D}^* \Phi, \Phi) \rangle_Q + \langle \mathsf{G}_1, \Phi_1 \rangle_\Omega - \langle \mathsf{G}_0, \Phi_0 \rangle_\Omega \,,$$

which allows us to write

$$\mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1) = \inf_{\mu \in \mathcal{M}(Q,\mathbb{X})} \sup_{\Phi \in C^1(Q,\mathbb{S}^n)} \mathcal{L}(\mu, \Phi)\,.$$

By changing the order of sup and inf, a formal calculation via integration by parts gives the dual problem:

$$\begin{aligned} \mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1) &\ge \sup_\Phi \inf_\mu \mathcal{L}(\mu, \Phi) \\ &= \sup_\Phi \left\{ \langle \mathsf{G}_1, \Phi_1 \rangle_\Omega - \langle \mathsf{G}_0, \Phi_0 \rangle_\Omega \,;\, \partial_t \Phi + \frac{1}{2}(\mathsf{D}^*\Phi)\Lambda_1^2(\mathsf{D}^*\Phi)^{\mathrm{T}} + \frac{1}{2}\Phi\Lambda_2^2\Phi \preceq 0 \right\}. \end{aligned} \tag{4.1}$$

We next use the Fenchel-Rockafellar theorem (Lemma 2.5) to show that the duality gap is zero, which will also give the existence of the minimiser to ($\mathcal{P}$) and the optimality conditions. For this, we define

$$C(Q, \mathcal{O}_\Lambda) := \{\varphi \in C(Q, \mathbb{X})\,;\, \varphi(x) \in \mathcal{O}_\Lambda\,,\ \forall x \in Q\}\,, \tag{4.2}$$

with $\mathcal{O}_\Lambda$ given in (3.1), which is a closed convex subset of $C(Q, \mathbb{X})$. We then define lower semicontinuous convex functions: $f(\Phi) = \langle \mathsf{G}_1, \Phi_1 \rangle_\Omega - \langle \mathsf{G}_0, \Phi_0 \rangle_\Omega$ for $\Phi \in C^1(Q, \mathbb{S}^n)$ and $g(\Xi) = \iota_{C(Q,\mathcal{O}_\Lambda)}(\Xi)$ for $\Xi \in C(Q, \mathbb{X})$. We also introduce the bounded linear operator: $L : \Phi \in C^1(Q, \mathbb{S}^n) \to (\partial_t \Phi, \mathsf{D}^*\Phi, \Phi) \in C(Q, \mathbb{X})$ with the dual operator L^*. These notions help us to write (4.1) as $\sup\{f(\Phi) - g(L\Phi)\,;\, \Phi \in C^1(Q, \mathbb{S}^n)\}$.

We now verify the condition in Lemma 2.5. We consider $\Phi = -\varepsilon t I + \frac{\varepsilon}{2} I \in C^1(Q, \mathbb{S}^n)$. It is clear that $f(\Phi)$ is finite and $L\Phi = (-\varepsilon I, 0, -\varepsilon t I + \frac{\varepsilon}{2} I)$ by $\mathsf{D}^*(I) = 0$. By a simple calculation, we have

$$\partial_t \Phi + \frac{1}{2}(\mathsf{D}^*\Phi)\Lambda_1^2(\mathsf{D}^*\Phi)^{\mathrm{T}} + \frac{1}{2}\Phi\Lambda_2^2\Phi = -\varepsilon I + \frac{1}{2}\varepsilon^2\left(-t + \frac{1}{2}\right)^2 \Lambda_2^2 \preceq -\varepsilon I + \frac{1}{8}\varepsilon^2 \Lambda_2^2\,,$$

which implies that for small enough ε and any $(t, x) \in Q$, $(L\Phi)(t, x)$ is in the interior of $\mathcal{O}_\Lambda$ and hence g is continuous at $L\Phi$. Then Lemma 2.5 readily gives

$$\min_{\mu \in \mathcal{M}(Q,\mathbb{X})} f^*(L^*\mu) + g^*(\mu) = \sup_{\Phi \in C^1(Q,\mathbb{S}^n)} f(\Phi) - g(L\Phi)\,, \tag{4.3}$$

where $f^*(L^*\mu) = \sup\{\langle \mu, L\Phi \rangle_Q - f(\Phi)\,;\, \Phi \in C^1(Q, \mathbb{S}^n)\}$ can be easily computed as $\iota_{\mathcal{CE}([0,1];\mathsf{G}_0,\mathsf{G}_1)}$ by linearity of f, while $g^*(\mu)$ is nothing else than $\mathcal{J}_{\Lambda,Q}(\mu)$ by the following lemma, which is a direct application of general results [13, 83]. We sketch the proof in Appendix A for completeness.

Lemma 4.1. *Let $\mathcal{X}$ be a compact separable metric space and $C(\mathcal{X}, \mathcal{O}_\Lambda)$ be defined in (4.2). Then, we have*

$$\iota^*_{C(\mathcal{X},\mathcal{O}_\Lambda)} = \sup_{\Xi \in L^\infty_{|\mu|}(\mathcal{X},\mathcal{O}_\Lambda)} \langle \mu, \Xi \rangle_{\mathcal{X}} = \mathcal{J}_{\Lambda,\mathcal{X}}(\mu)\,, \quad \textit{for } \mu \in \mathcal{M}(\mathcal{X}, \mathbb{X})\,, \tag{4.4}$$

which is proper convex and lower semicontinuous with respect to the weak topology of $\mathcal{M}(\mathcal{X}, \mathbb{X})$. Moreover, the subgradient $\partial \mathcal{J}_{\Lambda,\mathcal{X}}(\mu)$ in $C(\mathcal{X}, \mathbb{X})$ is given as follows:*

$$\partial \mathcal{J}_{\Lambda,\mathcal{X}}(\mu)|_{C(\mathcal{X},\mathbb{X})} = \{\Xi \in C(\mathcal{X}, \mathcal{O}_\Lambda)\,;\, \Xi(x) \in \partial J_\Lambda(\mu_\lambda)(x)\,,\ \lambda\text{-a.e.}\}\,, \tag{4.5}$$

which is independent of the choice of the reference measure λ such that $|\mu| \ll \lambda$.

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

By the above arguments, we have shown the following result.

Theorem 4.2. *The optimisation problem* ($\mathcal{P}$) *always admits a minimiser* $\mu \in \mathcal{CE}([0,1]; \mathsf{G}_0, \mathsf{G}_1)$ *and a dual formulation with zero duality gap:*

$$\mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1) = \sup_{\Phi \in C^1(Q,\mathbb{S}^n)} \left\{ \langle \mathsf{G}_1, \Phi_1 \rangle_\Omega - \langle \mathsf{G}_0, \Phi_0 \rangle_\Omega - \iota_{C(Q,\mathcal{O}_\Lambda)}(\partial_t \Phi, \mathsf{D}^*\Phi, \Phi) \right\}, \tag{4.6}$$

where the sup *is attained at* $\Phi \in C^1(Q, \mathbb{S}^n)$ *if and only if there exists* $\mu = (\mathsf{G}, \mathsf{q}, \mathsf{R}) \in \mathcal{CE}([0,1]; \mathsf{G}_0, \mathsf{G}_1)$ *such that*

$$q_\lambda = G_\lambda (\mathsf{D}^*\Phi)\Lambda_1^2\,, \quad R_\lambda = G_\lambda \Phi \Lambda_2^2\,, \tag{4.7}$$

and

$$G_\lambda \cdot \left(\partial_t \Phi + \frac{1}{2}(\mathsf{D}^*\Phi)\Lambda_1^2(\mathsf{D}^*\Phi)^{\mathrm{T}} + \frac{1}{2}\Phi\Lambda_2^2\Phi \right) = 0\,, \tag{4.8}$$

for λ*-a.e.* $(t,x) \in Q$*. In this case,* μ *is also the minimiser to the problem* ($\mathcal{P}$).

As a consequence of Lemma 4.1 and the dual formulation (4.6), we have the sublinearity and the weak* lower semicontinuity of $\mathrm{WB}^2_\Lambda(\cdot,\cdot)$.

Corollary 4.3. $\mathrm{WB}^2_\Lambda(\cdot,\cdot)$ *is sublinear: for* $\alpha > 0$, $\mathsf{G}_0, \mathsf{G}_1, \widetilde{\mathsf{G}}_0, \widetilde{\mathsf{G}}_1 \in \mathcal{M}(\Omega, \mathbb{S}^n_+)$*, there holds*

$$\mathrm{WB}^2_\Lambda\big(\alpha\mathsf{G}_0, \alpha\mathsf{G}_1\big) = \alpha \mathrm{WB}^2_\Lambda\big(\mathsf{G}_0, \mathsf{G}_1\big)\,, \quad \mathrm{WB}^2_\Lambda\big(\mathsf{G}_0 + \widetilde{\mathsf{G}}_0, \mathsf{G}_1 + \widetilde{\mathsf{G}}_1\big) \le \mathrm{WB}^2_\Lambda\big(\mathsf{G}_0, \mathsf{G}_1\big) + \mathrm{WB}^2_\Lambda\big(\widetilde{\mathsf{G}}_0, \widetilde{\mathsf{G}}_1\big)\,. \tag{4.9}$$

Moreover, WB_Λ *is lower semicontinuous with respect to the weak* topology, that is, for any sequences* $\{\mathsf{G}_0^n\}_{n\in\mathbb{N}}$ *and* $\{\mathsf{G}_1^n\}_{n\in\mathbb{N}}$ *in* $\mathcal{M}(\Omega, \mathbb{S}^n_+)$ *that weak* converge to measures* $\mathsf{G}_0, \mathsf{G}_1 \in \mathcal{M}(\Omega, \mathbb{S}^n_+)$*, respectively, there holds*

$$\mathrm{WB}_\Lambda(\mathsf{G}_0, \mathsf{G}_1) \le \liminf_{n\to 0} \mathrm{WB}_\Lambda(\mathsf{G}_0^n, \mathsf{G}_1^n)\,. \tag{4.10}$$

Proof. Noting that $\mathcal{J}_{\Lambda,Q}(\mu)$ is positively homogeneous and convex, and hence sublinear, the sublinearity of $\mathrm{WB}^2_\Lambda(\cdot,\cdot)$ follows from definition ($\mathcal{P}$) and the linearity of the continuity equation. For the weak* lower semicontinuity, by (4.6), for any $\Phi \in C^1(Q,\mathbb{S}^n)$ with $\iota_{C(Q,\mathcal{O}_\Lambda)}(\partial_t\Phi, \mathsf{D}^*\Phi, \Phi) = 0$, there holds

$$\liminf_{n\to\infty} \mathrm{WB}^2_\Lambda(\mathsf{G}_0^n, \mathsf{G}_1^n) \ge \liminf_{n\to\infty} \langle \mathsf{G}_1^n, \Phi_1\rangle_\Omega - \langle \mathsf{G}_0^n, \Phi_0\rangle_\Omega = \langle \mathsf{G}_1, \Phi_1\rangle_\Omega - \langle \mathsf{G}_0, \Phi_0\rangle_\Omega\,, \tag{4.11}$$

by the weak* convergence of G_0^n and G_1^n. Then (4.10) follows by taking the sup of (4.11) over admissible Φ. □

In addition, we have the following explicit characterisation of the minimiser (i.e., geodesic; see Corollary 5.7) to ($\mathcal{P}$) for inflating measures from optimality conditions (4.7) and (4.8), which extends [16, Theorem 5] with a much simpler argument. For $\mathsf{G} \in \mathcal{M}(\Omega, \mathbb{S}^n_+)$ and $A \in \mathbb{S}^n_+$, we denote by G^A the inflating measure $A\mathsf{G}A \in \mathcal{M}(\Omega, \mathbb{S}^n_+)$.

Proposition 4.4. *For* $\mathsf{G} \in \mathcal{M}(\Omega, \mathbb{S}^n_+)$ *and matrices* $A_0, A_1 \in \mathbb{S}^n_+$*, we have*

$$\mathrm{WB}^2_\Lambda\big(\mathsf{G}^{A_0}, \mathsf{G}^{A_1}\big) = 2Tr\big(\Lambda_2^{-1}(A_1 - A_0)\mathsf{G}(\Omega)(A_1 - A_0)\Lambda_2^{-1}\big)\,, \tag{4.12}$$

with the minimiser $(\mathsf{G}_*, \mathsf{q}_*, \mathsf{R}_*) := (\mathsf{G}^{A_t}, 0, 2A_t\mathsf{G}(A_1 - A_0)) \in \mathcal{M}(Q, \mathbb{X})$*, where* $A_t := tA_1 + (1-t)A_0$ *for* $t \in [0,1]$.

Proof. Let us first assume that A_0 and A_1 are invertible. By a direct calculation, we have

$$\partial_t \mathsf{G}^{A_t} = (A_1 - A_0)\mathsf{G}A_t + A_t\mathsf{G}(A_1 - A_0)\,.$$

We define $\Phi = 2A_t^{-1}(A_1 - A_0)\Lambda_2^{-2}$ and find $\mathsf{R}_* = \mathsf{G}^{A_t}\Phi\Lambda_2^2$. It is also easy to see that $(\mathsf{G}_*, \mathsf{q}_*, \mathsf{R}_*)$ defined above is in the set $\mathcal{CE}\big([0,1]; \mathsf{G}^{A_0}, \mathsf{G}^{A_1}\big)$. Moreover, recalling $((A+\varepsilon H)^{-1} - A^{-1})/\varepsilon \to -A^{-1}HA^{-1}$ as $\varepsilon \to 0$ for invertible A and $H \in \mathbb{M}^n$ [9], we have

$$\partial_t \Phi = -2A_t^{-1}(A_1 - A_0)A_t^{-1}(A_1 - A_0)\Lambda_2^{-2} = -\Phi\Lambda_2^2\Phi/2\,.$$

By the above computations, we have verified the optimality conditions (4.7) and (4.8), which means that the measure $(\mathsf{G}_*, \mathsf{q}_*, \mathsf{R}_*)$ is the desired minimiser. Then, we can further compute

$$\mathrm{WB}_\Lambda^2\left(\mathsf{G}^{A_0}, \mathsf{G}^{A_1}\right) = \frac{1}{2}\int_0^1 \int_\Omega (\Phi\Lambda_2)\cdot \mathrm{d}\mathsf{G}^{A_t}(\Phi\Lambda_2)\,\mathrm{d}t = 2((A_1 - A_0)\Lambda_2^{-1})\cdot \mathsf{G}(\Omega)(A_1 - A_0)\Lambda_2^{-1}\,.$$

For general $A_0, A_1 \in \mathbb{S}_+^n$, we first see that $\mu_* := (\mathsf{G}^{A_t}, 0, 2A_t\mathsf{G}(A_1 - A_0))$ as above still satisfies the continuity equation and its associated action functional $\mathcal{J}_{\Lambda,Q}(\mu_*)$ gives the right-hand side of (4.12) by $\mathrm{Ran}(A_1 - A_0) \subset \mathrm{Ran}(A_t)$, which also means $\mathrm{WB}_\Lambda^2(\mathsf{G}^{A_0}, \mathsf{G}^{A_1}) \le \mathcal{J}_{\Lambda,Q}(\mu_*)$. To finish the proof, it suffices to show that the equality holds. For this, we consider $A_i^\varepsilon = A_i + \varepsilon I \in \mathbb{S}_{++}^n$ for $i = 0, 1$. Then, by triangle inequality of WB_Λ (see Proposition 5.2 below) and Lemma 3.9, we have $\mathrm{WB}_\Lambda(\mathsf{G}^{A_0^\varepsilon}, \mathsf{G}^{A_1^\varepsilon}) \to \mathrm{WB}_\Lambda(\mathsf{G}^{A_0}, \mathsf{G}^{A_1})$ as $\varepsilon \to 0$. The proof is completed by

$$\begin{aligned}\mathrm{WB}_\Lambda^2\left(\mathsf{G}^{A_0^\varepsilon}, \mathsf{G}^{A_1^\varepsilon}\right) &= 2Tr\left(\Lambda_2^{-1}(A_1^\varepsilon - A_0^\varepsilon)\mathsf{G}(\Omega)(A_1^\varepsilon - A_0^\varepsilon)\Lambda_2^{-1}\right)\\ &\to 2Tr\left(\Lambda_2^{-1}(A_1 - A_0)\mathsf{G}(\Omega)(A_1 - A_0)\Lambda_2^{-1}\right) = \mathcal{J}_{\Lambda,Q}(\mu_*)\,, \quad \varepsilon \to 0\,.\end{aligned}$$

□

4.2. Primal-dual formulations

We proceed to study in more depth the optimality conditions by viewing G as the main variable and (q,R) as the control variable, which will be useful in Section 5. We first observe

$$\mathrm{WB}_\Lambda^2(\mathsf{G}_0, \mathsf{G}_1) = \inf_{\mathsf{G}}\inf_{\mathsf{q},\mathsf{R}}\left\{\mathcal{J}_{\Lambda,Q}(\mu)\,;\,\mu = (\mathsf{G}, \mathsf{q}, \mathsf{R}) \in \mathcal{CE}_\infty([0, 1]; \mathsf{G}_0, \mathsf{G}_1)\right\}\,, \tag{4.13}$$

by taking the inf in $(\mathcal{P})$ over G and (q, R) separately. Recall the formulation (3.24) of $\mathcal{J}_{\Lambda,Q}(\mu)$, which motivates us to introduce a weighted semi-inner product:

$$\left\langle (u, W), (u', W')\right\rangle_{L^2_{\mathsf{G},\Lambda}(Q)} := \left\langle u\Lambda_1^\dagger, u'\Lambda_1^\dagger\right\rangle_{L^2_{\mathsf{G}}(Q)} + \left\langle W\Lambda_2^{-1}, W'\Lambda_2^{-1}\right\rangle_{L^2_{\mathsf{G}}(Q)}\,, \tag{4.14}$$

and the associated seminorm $\|\cdot\|_{L^2_{\mathsf{G},\Lambda}(Q)}$ on the space of measurable functions valued in $\mathbb{R}^{n\times k} \times \mathbb{M}^n$. The corresponding Hilbert space, denoted by $L^2_{\mathsf{G},\Lambda}(Q, \mathbb{R}^{n\times k} \times \mathbb{M}^n)$, is defined as the quotient space by the subspace $\mathrm{Ker}\left(\|\cdot\|_{L^2_{\mathsf{G},\Lambda}(Q)}\right)$. Hence, we can rewrite (3.24) as $\mathcal{J}_{\Lambda,Q}(\mu) = \|(G^\dagger q, G^\dagger R)\|^2_{L^2_{\mathsf{G},\Lambda}(Q)}/2$. Moreover, we define the set

$$\mathcal{AC}([0, 1]; \mathsf{G}_0, \mathsf{G}_1) := \{\mathsf{G} \in \mathcal{M}(Q, \mathbb{S}^n)\,; \exists(\mathsf{q},\mathsf{R}) \in \mathcal{M}(Q, \mathbb{R}^{n\times k} \times \mathbb{M}^n)\ \text{s.t.}\ (\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{CE}_\infty([0, 1]; \mathsf{G}_0, \mathsf{G}_1)\}\,, \tag{4.15}$$

and the associated energy functional: for $\mathsf{G} \in \mathcal{AC}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$,

$$\mathcal{J}^\Lambda_{\mathsf{G}_0,\mathsf{G}_1}(\mathsf{G}) := \inf_{(\mathsf{q},\mathsf{R})}\left\{\frac{1}{2}\|(G^\dagger q, G^\dagger R)\|^2_{L^2_{\mathsf{G},\Lambda}(Q)}\,;\,(\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{CE}_\infty([0, 1]; \mathsf{G}_0, \mathsf{G}_1)\right\}. \tag{4.16}$$

We will see in Remark 5.6 that $\mathcal{AC}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$ is closely related to the set of absolutely continuous curves in the metric space $(\mathcal{M}(\Omega, \mathbb{S}_+^n), \mathrm{WB}_\Lambda)$. With the help of these notions, (4.13) can be reformulated in a compact form:

$$\mathrm{WB}_\Lambda^2(\mathsf{G}_0, \mathsf{G}_1) = \inf_{\mathsf{G}\in\mathcal{AC}([0,1];\,\mathsf{G}_0,\mathsf{G}_1)} \mathcal{J}^\Lambda_{\mathsf{G}_0,\mathsf{G}_1}(\mathsf{G})\,. \tag{4.17}$$

Similarly to (3.24), by Lemma 3.3, we also note that for $(\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{CE}_\infty([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$, the weak formulation (3.13) can be written as

$$\left\langle\left(\mathsf{D}^*\Phi\Lambda_1^2, \Phi\Lambda_2^2\right), \left(G^\dagger q, G^\dagger R\right)\right\rangle_{L^2_{\mathsf{G},\Lambda}(Q)} = l_{\mathsf{G}}(\Phi)\,, \quad \forall\Phi \in C^1(Q, \mathbb{S}^n)\,, \tag{4.18}$$

where $l_{\mathsf{G}}(\,\cdot\,)$ for $\mathsf{G} \in \mathcal{AC}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$ is a linear functional on $C^1(Q, \mathbb{S}^n)$ defined by

$$l_{\mathsf{G}}(\Phi) = \langle \mathsf{G}_1, \Phi_1\rangle_\Omega - \langle \mathsf{G}_0, \Phi_0\rangle_\Omega - \langle \mathsf{G}, \partial_t\Phi\rangle_Q\,. \tag{4.19}$$

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

Define an injective map $\Pi : \Phi \to (\mathsf{D}^*\Phi\Lambda_1^2, \Phi\Lambda_2^2)$ for $\Phi \in C^1(Q, \mathbb{S}^n)$ and denote $\widetilde{l}_{\mathsf{G}} := l_{\mathsf{G}} \circ \Pi^{-1}$ on the image of Π. In view of (4.18), the functional $\widetilde{l}_{\mathsf{G}}$ can be uniquely extended to the space

$$H_{\mathsf{G},\Lambda}(\mathsf{D}^*) := \overline{\{\Pi(\Phi)\,;\, \Phi \in C^1(Q, \mathbb{S}^n)\}}^{\|\cdot\|_{L^2_{\mathsf{G},\Lambda}(Q)}}\,, \tag{4.20}$$

with the norm estimate

$$\|\widetilde{l}_{\mathsf{G}}\|_{H^*_{\mathsf{G},\Lambda}(\mathsf{D}^*)} \le \|(G^\dagger q, G^\dagger R)\|_{L^2_{\mathsf{G},\Lambda}(Q)}\,. \tag{4.21}$$

We emphasise that such an extension is independent of the choice of (q,R) that satisfies $(\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{CE}_\infty([0,1]; \mathsf{G}_0, \mathsf{G}_1)$.

Next, we show that (4.16) admits a unique minimiser (q,R) that satisfies the equality in (4.21). Note that (u, W) and $(u\mathbb{P}_{\Lambda_1}, W\mathbb{P}_{\Lambda_2})$ are equivalent in $L^2_{\mathsf{G},\Lambda}(Q, \mathbb{R}^{n\times k} \times \mathbb{M}^n)$, where $\mathbb{P}_{\Lambda_i}$ is the orthogonal projection to $\mathrm{Ran}(\Lambda_i)$. Hence, for any $(u, W) \in L^2_{\mathsf{G},\Lambda}(Q, \mathbb{R}^{n\times k} \times \mathbb{M}^n)$, we can assume $\mathrm{Ran}(u^{\mathrm{T}}) \subset \mathrm{Ran}(\Lambda_1)$ and $\mathrm{Ran}(W^{\mathrm{T}}) \subset \mathrm{Ran}(\Lambda_2)$. Then, it holds that any $L^2_{\mathsf{G},\Lambda}$-field (u, W) satisfying $\langle(\mathsf{D}^*\Phi\Lambda_1^2, \Phi\Lambda_2^2), (u, W)\rangle_{L^2_{\mathsf{G},\Lambda}(Q)} = l_{\mathsf{G}}(\Phi)$, $\forall \Phi \in C^1(Q, \mathbb{S}^n)$, induces a measure $(\mathsf{q},\mathsf{R}) := (\mathsf{G}u, \mathsf{G}W)$ such that $(\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{CE}_\infty([0,1]; \mathsf{G}_0, \mathsf{G}_1)$. This observation implies that $\mathcal{J}^\Lambda_{\mathsf{G}_0,\mathsf{G}_1}(\mathsf{G})$ is actually a uniquely solvable minimum norm problem with an affine constraint:

$$\begin{aligned}\mathcal{J}^\Lambda_{\mathsf{G}_0,\mathsf{G}_1}(\mathsf{G}) = \inf \Big\{ \frac{1}{2}\|(u, W)\|^2_{L^2_{\mathsf{G},\Lambda}(Q)}\,;\, (u, W) \in L^2_{\mathsf{G},\Lambda}(Q, \mathbb{R}^{n\times k} \times \mathbb{M}^n) \text{ such that} \\ \big\langle(\mathsf{D}^*\Phi\Lambda_1^2, \Phi\Lambda_2^2), (u, W)\big\rangle_{L^2_{\mathsf{G},\Lambda}(Q)} = l_{\mathsf{G}}(\Phi)\,,\ \forall \Phi \in C^1(Q, \mathbb{S}^n)\Big\}\,.\end{aligned} \tag{4.22}$$

The unique minimiser (u_*, W_*) to (4.22) is given by the orthogonal projection of 0 on the constraint set, equivalently, the Riesz representation of the functional $\widetilde{l}_{\mathsf{G}}$ on the space $H_{\mathsf{G},\Lambda}(\mathsf{D}^*)$. It then follows that $(\mathsf{q}_*, \mathsf{R}_*) := (\mathsf{G}u_*, \mathsf{G}W_*)$ is the desired minimiser to (4.16) and there holds

$$\|\widetilde{l}_{\mathsf{G}}\|_{H^*_{\mathsf{G},\Lambda}(\mathsf{D}^*)} = \|(u_*, W_*)\|_{L^2_{\mathsf{G},\Lambda}(Q)} = \|(G^\dagger q_*, G^\dagger R_*)\|_{L^2_{\mathsf{G},\Lambda}(Q)}\,. \tag{4.23}$$

We summarise the above facts in the following useful result.

Theorem 4.5. $\mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1)$ *has the following representation:*

$$\mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1) = \inf_{\mathsf{G}\in\mathcal{AC}([0,1];\,\mathsf{G}_0,\mathsf{G}_1)} \mathcal{J}^\Lambda_{\mathsf{G}_0,\mathsf{G}_1}(\mathsf{G}) \quad \textit{with} \quad \mathcal{J}^\Lambda_{\mathsf{G}_0,\mathsf{G}_1}(\mathsf{G}) = \frac{1}{2}\|(u_*, W_*)\|^2_{L^2_{\mathsf{G},\Lambda}(Q)}\,,$$

where (u_*, W_*) is the Riesz representation of $\widetilde{l}_{\mathsf{G}}$ in $H_{\mathsf{G},\Lambda}(\mathsf{D}^*)$ that uniquely solves the minimum norm problem (4.22).

Moreover, $\mathcal{J}^\Lambda_{\mathsf{G}_0,\mathsf{G}_1}(\mathsf{G})$ admits the following dual formulation:

$$\mathcal{J}^\Lambda_{\mathsf{G}_0,\mathsf{G}_1}(\mathsf{G}) = \sup\Big\{ l_{\mathsf{G}}(\Phi) - \frac{1}{2}\|(\mathsf{D}^*\Phi\Lambda_1^2, \Phi\Lambda_2^2)\|^2_{L^2_{\mathsf{G},\Lambda}(Q)}\,;\, \Phi \in C^1(Q, \mathbb{S}^n)\Big\}\,. \tag{4.24}$$

Proof. It suffices to derive the dual formulation (4.24) of $\mathcal{J}^\Lambda_{\mathsf{G}_0,\mathsf{G}_1}$. For this, we first note

$$\frac{1}{2}\|(u, W)\|^2_{L^2_{\mathsf{G},\Lambda}(Q)} = \sup_{(u',W')\in L^2_{\mathsf{G},\Lambda}(Q,\mathbb{R}^{n\times k}\times\mathbb{M}^n)} \langle(u, W), (u', W')\rangle_{L^2_{\mathsf{G},\Lambda}(Q)} - \frac{1}{2}\|(u', W')\|^2_{L^2_{\mathsf{G},\Lambda}(Q)}\,,$$

which further implies, by $(u_*, W_*) \in H_{\mathsf{G},\Lambda}(\mathsf{D}^*) \subset L^2_{\mathsf{G},\Lambda}(Q, \mathbb{R}^{n\times k} \times \mathbb{M}^n)$, for any $\Phi \in C^1(Q, \mathbb{S}^n)$,

$$\begin{aligned}\mathcal{J}^\Lambda_{\mathsf{G}_0,\mathsf{G}_1}(\mathsf{G}) = \frac{1}{2}\|(u_*, W_*)\|^2_{L^2_{\mathsf{G},\Lambda}(Q)} &\ge \langle(u_*, W_*), \big(\mathsf{D}^*\Phi\Lambda_1^2, \Phi\Lambda_2^2\big)\rangle_{L^2_{\mathsf{G},\Lambda}(Q)} - \frac{1}{2}\|(\mathsf{D}^*\Phi\Lambda_1^2, \Phi\Lambda_2^2)\|^2_{L^2_{\mathsf{G},\Lambda}(Q)} \\ &= l_G(\Phi) - \frac{1}{2}\|(\mathsf{D}^*\Phi\Lambda_1^2, \Phi\Lambda_2^2)\|^2_{L^2_{\mathsf{G},\Lambda}(Q)}\,.\end{aligned} \tag{4.25}$$

Then, recalling (4.20) and choosing a sequence $\{(\mathsf{D}^*\Phi_n\Lambda_1^2, \Phi_n\Lambda_2^2)\}$ with $\Phi_n \in C^1(Q, \mathbb{S}^n)$ in (4.25) that approximates (u_*, W_*) gives the desired (4.24). □

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

4.3. Varying weight matrices

We regard WB_Λ as a family of distances indexed by Λ and investigate the behaviours of WB_Λ and its minimiser when Λ varies, in particular, when $|\Lambda_1|$ or $|\Lambda_2|$ tends to zero or infinity. We give a partial answer to this question in the following proposition. For ease of exposition, we introduce

$$\mathcal{J}^q_{\Lambda_1}(\mu) = \mathcal{J}_{\Lambda,Q}((\mathsf{G},\mathsf{q},0)), \quad \mathcal{J}^R_{\Lambda_2}(\mu) = \mathcal{J}_{\Lambda,Q}((\mathsf{G},0,\mathsf{R})) \quad \text{for } \mu \in \mathcal{M}(Q,\mathbb{X}). \tag{4.26}$$

Proposition 4.6. *Let $\mathsf{G}_0, \mathsf{G}_1 \in \mathcal{M}(\Omega, \mathbb{S}^n_+)$ and $\mu_{*,\Lambda}$ denote the minimiser to $\mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1)$ ($\mathcal{P}$). It holds that $\mathrm{WB}^2_{(\Lambda_1,\Lambda_2)}(\mathsf{G}_0, \mathsf{G}_1) \to \mathrm{WB}^2_{(0,\Lambda_2)}(\mathsf{G}_0, \mathsf{G}_1)$ as $\|\Lambda_1\|_{\mathrm{F}} \to 0$, and for any sequence $\{\Lambda_{1,j}\}_{j\in\mathbb{N}} \subset \mathbb{S}^k_+$ with $\|\Lambda_{1,j}\|_{\mathrm{F}} \to 0$, the associated minimiser $\mu_{*,(\Lambda_{1,j},\Lambda_2)}$, up to a subsequence, weak* converges to a minimiser μ_* to $\mathrm{WB}^2_{(0,\Lambda_2)}(\mathsf{G}_0, \mathsf{G}_1)$.*

Proof. We first claim that $\|\Lambda_1\|^2_{\mathrm{F}} \mathcal{J}^q_{\Lambda_1}(\mu_{*,\Lambda})$ and $\mathcal{J}^R_{\Lambda_2}(\mu_{*,\Lambda})$ are bounded when $\|\Lambda_1\|_{\mathrm{F}} \to 0$, which, by estimates (3.25) and (3.28), implies that $\mu_{*,\Lambda}$ is bounded in $\mathcal{M}(Q,\mathbb{X})$. For this, we consider the set

$$\mathcal{CE}_{\Lambda_1,q} := \arg\min\{\mathcal{J}^q_{\Lambda_1}(\mu)\,;\, \mu \in \mathcal{CE}([0,1];\mathsf{G}_0,\mathsf{G}_1)\}. \tag{4.27}$$

Similarly to the proof of Lemma 3.9, we have that $\mathcal{CE}_{\Lambda_1,q}$ is nonempty and contains at least one element with $\mathsf{q} = 0$ and $\min\{\mathcal{J}^q_{\Lambda_1}(\mu)\,;\, \mu \in \mathcal{CE}([0,1];\mathsf{G}_0,\mathsf{G}_1)\} = 0$. Since $\mu_{*,\Lambda}$ minimises $\mathcal{J}_{\Lambda,Q}(\,\cdot\,)$, it follows that

$$\mathcal{J}_{\Lambda,Q}(\mu_{*,\Lambda}) = \mathcal{J}^q_{\Lambda_1}(\mu_{*,\Lambda}) + \mathcal{J}^R_{\Lambda_2}(\mu_{*,\Lambda}) \le \mathcal{J}_{\Lambda,Q}(\mu) = \mathcal{J}^R_{\Lambda_2}(\mu), \quad \forall \mu = (\mathsf{G},0,\mathsf{R}) \in \mathcal{CE}_{\Lambda_1,q}. \tag{4.28}$$

Noting $\{(\mathsf{G},0,\mathsf{R}) \in \mathcal{CE}_{\Lambda_1,q}\} = \{(\mathsf{G},0,\mathsf{R}) \in \mathcal{CE}([0,1];\mathsf{G}_0,\mathsf{G}_1)\}$, (4.28) yields that $\mathcal{J}^R_{\Lambda_2}(\mu_{*,\Lambda})$ is bounded by a constant independent of Λ_1. Moreover, multiplying $\|\Lambda_1\|^2_{\mathrm{F}}$ on both sides of (4.28) and then letting $\|\Lambda_1\|_{\mathrm{F}} \to 0$, we obtain

$$\lim_{\|\Lambda_1\|_{\mathrm{F}}\to 0} \|\Lambda_1\|^2_{\mathrm{F}}\, \mathcal{J}^q_{\Lambda_1}(\mu_{*,\Lambda}) = 0. \tag{4.29}$$

Then the boundedness of $\|\Lambda_1\|^2_{\mathrm{F}} \mathcal{J}^q_{\Lambda_1}(\mu_{*,\Lambda})$ for small enough $\|\Lambda_1\|_{\mathrm{F}}$ follows. We complete the proof of the claim.

By the boundedness of $\|\mu_{*,\Lambda}\|_{\mathrm{TV}}$ as $\|\Lambda_1\|_{\mathrm{F}} \to 0$, we are allowed to take a subsequence $\{\Lambda_{1,j}\}_{j\in\mathbb{N}}$ in $\mathbb{S}^n_+$ such that the minimiser $\mu_{*,\widetilde{\Lambda}_j}$ with $\widetilde{\Lambda}_j = (\Lambda_{1,j}, \Lambda_2)$ weak* converges to a measure $\mu_* \in \mathcal{M}(Q,\mathbb{X})$ when $n \to \infty$, which clearly satisfies $\mu_* \in \mathcal{CE}([0,1];\mathsf{G}_0,\mathsf{G}_1)$. Then, by the weak* lower semicontinuity of $\mathcal{J}^R_{\Lambda_2}$ and (4.28), we have

$$\mathcal{J}^R_{\Lambda_2}(\mu_*) \le \liminf_{j\to\infty} \mathcal{J}^R_{\Lambda_2}(\mu_{*,\widetilde{\Lambda}_j}) \le \limsup_{j\to\infty} \mathrm{WB}^2_{\widetilde{\Lambda}_j}(\mathsf{G}_0,\mathsf{G}_1) \le \inf\{\mathcal{J}^R_{\Lambda_2}(\mu)\,;\, \mu = (\mathsf{G},0,\mathsf{R}) \in \mathcal{CE}_{\Lambda_1,q}\}. \tag{4.30}$$

The right-hand side of (4.30) is recognised as $\mathrm{WB}_{(0,\Lambda_2)}(\mathsf{G}_0,\mathsf{G}_1)$ and the inf is attained; see Remark 3.11 and Theorem 4.2. Also, by (3.25) and (4.29), it holds that the limit measure $\mu_* \in \mathcal{CE}([0,1];\mathsf{G}_0,\mathsf{G}_1)$ is of the form $\mu_* = (\mathsf{G}_*, 0, \mathsf{R}_*)$. The proof is completed by (4.30). □

Proposition 4.6 above tells us that the measure q is forced to be nearly zero, if the transportation part is given too much weight (i.e., $\|\Lambda_1\|_{\mathrm{F}}$ is small, cf. (3.24)), equivalently, if the problem is on a large scale (cf. Remark 3.17). It is also possible and interesting to consider other limiting regimes, e.g., $\|\Lambda_1\|_{\mathrm{F}} \to \infty$, $\|\Lambda_2\|_{\mathrm{F}} \to 0$, or only let part of eigenvalues of Λ_i vanish, which, however, is beyond the scope of this work.

5. Geometric properties and Riemannian interpretation

In this section, we shall study the space $\mathcal{M}(\Omega,\mathbb{S}^n_+)$ equipped with the distance $\mathrm{WB}_\Lambda(\cdot,\cdot)$ from the metric point of view. In particular, we will prove that $(\mathcal{M}(\Omega,\mathbb{S}^n_+), \mathrm{WB}_\Lambda)$ is a complete geodesic space with a Riemannian interpretation. We first show that $\mathrm{WB}_\Lambda(\cdot,\cdot)$ is indeed a metric on $\mathcal{M}(\Omega,\mathbb{S}^n_+)$, which is

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

a simple corollary of the following characterisation of $\mathrm{WB}_\Lambda(\cdot,\cdot)$ by standard reparameterisation techniques (cf. [1, Lemma 1.1.4] or [34, Theorem 5.4]). We denote by $\widetilde{\mathcal{CE}}([a,b];\mathsf{G}_0,\mathsf{G}_1)$ the set of measures $\mu \in \mathcal{CE}([a,b];\mathsf{G}_0,\mathsf{G}_1)$ that can be disintegrated as $\mu = \int_a^b \delta_t \otimes \mu_t \,\mathrm{d}t$. It is clear that $\mathcal{CE}_\infty \subset \widetilde{\mathcal{CE}} \subset \mathcal{CE}$.

Lemma 5.1. *For $\mathsf{G}_0, \mathsf{G}_1 \in \mathcal{M}(\Omega, \mathbb{S}^n_+)$ and $b > a > 0$, there holds*

$$\mathrm{WB}_\Lambda(\mathsf{G}_0,\mathsf{G}_1) = \inf_{\mu\in\widetilde{\mathcal{CE}}([a,b];\mathsf{G}_0,\mathsf{G}_1)} \int_a^b \mathcal{J}_{\Lambda,\Omega}(\mu_t)^{1/2}\,\mathrm{d}t\,. \tag{5.1}$$

Moreover, the minimiser to the problem ($\mathcal{P}$) gives a constant-speed minimiser μ to (5.1), which satisfies

$$(b-a)J_{\Lambda,\Omega}(\mu_t)^{1/2} = \mathrm{WB}_\Lambda(\mathsf{G}_0,\mathsf{G}_1) \quad \textit{for a.e. } t \in [a,b]\,. \tag{5.2}$$

The proof is provided in Appendix A for completeness. The above lemma is an analogue of a well-known geometric fact that minimising the energy of a parametric curve is the same as minimising its length with constant-speed constraint [40]. The following result summarises some fundamental properties of $(\mathcal{M}(\Omega,\mathbb{S}^n_+),\mathrm{WB}_\Lambda)$.

Proposition 5.2. *$(\mathcal{M}(\Omega,\mathbb{S}^n_+),\mathrm{WB}_\Lambda)$ is a complete metric space. Moreover, the topology induced by the metric WB_Λ is stronger than the weak* one, i.e., $\lim_{n\to\infty}\mathrm{WB}_\Lambda(\mathsf{G}^n,\mathsf{G}) = 0$ implies the weak* convergence of G^n to G.*

Remark 5.3. We should emphasise that stronger in Proposition 5.2 above means at least as strong as. In the special case of WFR distance ($\mathcal{P}_{\mathrm{WFR}}$), one can show [65, Theorem 7.15] that $\mathrm{WFR}(\cdot,\cdot)$ metrizes the weak* topology on $\mathcal{M}(\Omega,\mathbb{R}_+)$. However, the exact characterisation of the topology induced by a general metric $\mathrm{WB}_\Lambda(\cdot,\cdot)$ is still open. In addition, given the multi-component nature of our matrix-valued transport problem, one can expect that there may be some interesting connections between our model $(\mathcal{M}(\Omega,\mathbb{S}^n_+),\mathrm{WB}_\Lambda)$ and the multimaterial transport problem [11, 70], which deals with the simultaneous transportation of vector-valued measures along a network or graph and can exhibit the branching behaviour. The detailed investigation of these problems is beyond the scope of this work and left for future work.

The proof of Proposition 5.2 needs a priori estimates (3.25) and (3.28), and the following lemma, which is a direct consequence of Lemma 3.9.

Lemma 5.4. *A subset of $\mathcal{M}(\Omega,\mathbb{S}^n_+)$ is bounded with respect to the distance WB_Λ if and only if it is bounded with respect to the total variation norm. Hence, a bounded set in $(\mathcal{M}(\Omega,\mathbb{S}^n_+),\mathrm{WB}_\Lambda)$ is weak* relatively compact.*

Proof of Proposition 5.2. First, note that WB_Λ is a function from $\mathcal{M}(\Omega,\mathbb{S}^n_+)\times\mathcal{M}(\Omega,\mathbb{S}^n_+)$ to $[0,+\infty)$. It is also easy to check $\mathrm{WB}_\Lambda(\mathsf{G}_0,\mathsf{G}_1)=0$ for $\mathsf{G}_0=\mathsf{G}_1$ by considering the constant curve $\mathsf{G}_t=\mathsf{G}_0$ with $\mathsf{q}=\mathsf{R}=0$, the symmetry $\mathrm{WB}_\Lambda(\mathsf{G}_0,\mathsf{G}_1)=\mathrm{WB}_\Lambda(\mathsf{G}_1,\mathsf{G}_0)$ by Lemma 3.15 and the triangle inequality by (5.1). Then, to show that WB_Λ is a metric, it suffices to prove that $\mathrm{WB}_\Lambda(\mathsf{G}_0,\mathsf{G}_1)=0$ implies $\mathsf{G}_0=\mathsf{G}_1$.

For this, suppose that $\mu=(\mathsf{G},\mathsf{q},\mathsf{R})$ is a minimiser to ($\mathcal{P}$) with $\mathcal{J}_{\Lambda,Q}(\mu)=0$. Recalling the formula (3.24), we have $(\mathsf{q},\mathsf{R})=0$. Then, taking test functions $\Phi(t,x)=\Psi(x)$ with $\Psi(x)\in C^1(\Omega,\mathbb{S}^n)$ in (3.13), we find $\langle \mathsf{G}_1-\mathsf{G}_0,\Psi\rangle_\Omega=0$, $\forall\Psi\in C^1(\Omega,\mathbb{S}^n)$, which implies $\mathsf{G}_0=\mathsf{G}_1$. Next, we show that the metric space $(\mathcal{M}(\Omega,\mathbb{S}^n_+),\mathrm{WB}_\Lambda)$ is complete. Let $\{\mathsf{G}^n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $(\mathcal{M}(\Omega,\mathbb{S}^n_+),\mathrm{WB}_\Lambda)$, and hence also bounded in WB_Λ. By Lemma 5.4, we have that G^n, up to a subsequence, weak* converges to a measure $\mathsf{G}\in\mathcal{M}(\Omega,\mathbb{S}^n_+)$. Then, by Corollary 4.3 and the fact that $\{\mathsf{G}^n\}$ is a Cauchy sequence, for small $\varepsilon>0$ and large enough m, there holds

$$\varepsilon \ge \liminf_{n\to 0}\mathrm{WB}_\Lambda(\mathsf{G}^n,\mathsf{G}^m) \ge \mathrm{WB}_\Lambda(\mathsf{G},\mathsf{G}^m)\,,$$

which immediately gives $\mathrm{WB}_\Lambda(\mathsf{G},\mathsf{G}^m)\to 0$ as $m\to\infty$. To finish, we show that G^n weak* converges to G if G^n converges to G in $(\mathcal{M}(\Omega,\mathbb{S}^n_+),\mathrm{WB}_\Lambda)$. To do so, it suffices to note that by a similar argument as above, every subsequence of G^n has a weak* convergent sub-subsequence to G, which readily gives the weak* convergence of G^n to G. □

The main aim of this section is to show that $(\mathcal{M}(\Omega, \mathbb{S}^n), \mathrm{WB}_\Lambda)$ is a geodesic space and then equip it with some differential structure that is consistent with the metric structure, in the spirit of [1, 34].

For the reader's convenience, we recall some basic concepts for the analysis in metric spaces [2]. Let (X, d) be a metric space and $\{\omega_t\}_{t\in[a,b]}$ be a curve in (X, d) (i.e., a continuous map from $[a, b]$ to X). We say that it is absolutely continuous if there exists a L^1-function g such that $d(\omega_t, \omega_s) \le \int_s^t g(r)\,\mathrm{d}r$ for any $a \le s \le t \le b$. Moreover, the curve is said to have finite p-energy if $g \in L^p([a, b], \mathbb{R})$.

The metric derivative $|\omega_t'|$ of $\{\omega_t\}_{t\in[a,b]}$ at the time point t is defined by $|\omega_t'| := \lim_{\delta\to 0} |\delta|^{-1} d(\omega_{t+\delta}, \omega_t)$, if the limit exists. It can be shown [1, Theorem 1.1.2] that for an absolutely continuous curve ω_t, the metric derivative $|\omega_t'|$ is well-defined for a.e. $t \in [a, b]$ and satisfies $|\omega_t'| \le g(t)$.

The length $\mathrm{L}(\omega_t)$ of an absolutely continuous curve $\{\omega_t\}_{t\in[a,b]}$ is defined as $\mathrm{L}(\omega_t) = \int_a^b |\omega_t'|\,\mathrm{d}t$, which is invariant with respect to the reparameterisation. Then, (X, d) is a geodesic space if for any $x, y \in X$, there holds

$$d(x, y) = \min\{\mathrm{L}(\omega_t);\ \{\omega_t\}_{t\in[0,1]} \text{ is absolutely continuous with } \omega(0) = x\,,\ \omega(1) = y\}, \tag{5.3}$$

where the minimiser exists and is called the (minimizing) geodesic between x and y. Recall [1, Lemma 1.1.4] that any absolutely continuous curve can be reparameterised as a Lipschitz one with constant metric derivative $|\omega_t'| = \mathrm{L}(\omega_t)$ a.e.. Hence, we can always assume that the geodesic is constant-speed (i.e., $|\omega_t'|$ is constant a.e.). Then, it is clear from definition (5.3) that a curve $\{\omega_t\}_{t\in[0,1]}$ is a constant-speed geodesic if and only if it satisfies $d(\omega_s, \omega_t) = |t - s| d(\omega_0, \omega_1)$ for any $0 < s < t < 1$.

From the above concepts, we see that for our purpose, a key step is to characterise the absolutely continuous curves in the metric space $(\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda)$, which is given by the following theorem extended from [34, Theorem 5.17].

Theorem 5.5. *A curve $\{\mathsf{G}_t\}_{t\in[a,b]}$, $b > a > 0$, is absolutely continuous with respect to the metric WB_Λ if and only if there exists $(\mathsf{q}, \mathsf{R}) \in \mathcal{M}(Q, \mathbb{R}^{n\times k} \times \mathbb{M}^n)$ such that $\mu = (\mathsf{G}, \mathsf{q}, \mathsf{R}) \in \widetilde{\mathcal{CE}}([a, b]; \mathsf{G}_0, \mathsf{G}_1)$ and*

$$\int_a^b \mathcal{J}_{\Lambda,\Omega}(\mu_t)^{1/2}\,\mathrm{d}t < +\infty\,. \tag{5.4}$$

In this case, the metric derivative $|\mathsf{G}_t'|$ satisfies

$$|\mathsf{G}_t'| \le \mathcal{J}_{\Lambda,\Omega}(\mu_t)^{1/2} \quad \text{for a.e. } t \in [a, b], \tag{5.5}$$

and there exists unique $(\mathsf{q}_*, \mathsf{R}_*)$ such that the equality in (5.5) holds a.e., where the uniqueness is in the sense of equivalence class: $(\mathsf{q},\mathsf{R}) \sim (\mathsf{q}', \mathsf{R}')$ if and only if $\mathcal{J}_{\Lambda,Q_a^b}((\mathsf{G}, \mathsf{q}\text{-}\mathsf{q}', \mathsf{R}\text{-}\mathsf{R}')) = 0$. If G_t has finite 2-energy, then $(\mathsf{q}_*, \mathsf{R}_*) = (\mathsf{G}u_*, \mathsf{G}W_*)$ with the $L^2_{\mathsf{G},\Lambda}$-field (u_*, W_*) given in Theorem 4.5.

Remark 5.6. As a corollary of Theorem 5.5, we have that $\mathcal{AC}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$ in (4.15) is nothing else than the set of absolutely continuous curves with finite 2-energy.

Proof. It suffices to consider the case $[a, b] = [0, 1]$. We first consider the trivial *if* part. For $\mu \in \widetilde{\mathcal{CE}}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$ with the property (5.4), it follows from (5.1) that

$$\mathrm{WB}_\Lambda(\mathsf{G}_s, \mathsf{G}_t) \le \int_s^t \mathcal{J}_{\Lambda,\Omega}(\mu_\tau)^{1/2}\,\mathrm{d}\tau \quad \forall 0 \le s \le t \le 1\,,$$

which, by definition, readily implies that $\{\mathsf{G}_t\}_{t\in[0,1]}$ is absolutely continuous and (5.5) holds. We now consider the *only if* part. Let $\{\mathsf{G}_t\}_{t\in[0,1]}$ be an absolutely continuous curve, which, by reparameterisation, can be further assumed to be Lipschitz with the Lipschitz constant denoted by $\mathrm{Lip}(\mathsf{G}_t)$. We will approximate it by piecewise constant-speed curves. We fix an integer $N \in \mathbb{N}$ with the step size $\tau = 2^{-N}$. Let $\{\mu_t^{k,N}\}_{t\in[(k-1)\tau,k\tau]}$ be a minimiser to $(\mathcal{P}')$ with $[a, b] = [(k-1)\tau, k\tau]$, which satisfies

$$\tau^{1/2}\mathcal{J}_{\Lambda,\Omega}(\mu_t^{k,N})^{1/2} = \tau^{-1/2}\mathrm{WB}_\Lambda(\mathsf{G}_{(k-1)\tau}, \mathsf{G}_{k\tau}) \le \left(\int_{(k-1)\tau}^{k\tau} |\mathsf{G}_t'|^2\,\mathrm{d}t\right)^{1/2}, \quad a.e.\ t \in [(k-1)\tau, k\tau]\,, \tag{5.6}$$

by Lemma 5.1 and the absolute continuity of G_t. We glue the curves $\{\mu_t^{k,N}\}_{t\in[(k-1)\tau,k\tau]}$ with $k=1,\ldots,2^N$ and obtain a new one $\{\mu_t^N=(\mathsf{G}_t^N,\mathsf{q}_t^N,\mathsf{R}_t^N)\}_{t\in[0,1]}\in\mathcal{CE}_\infty([0,1];\mathsf{G}_0,\mathsf{G}_1)$.

Next, note that for any $(a,b)\subset[0,1]$, there exists $k_1^N,k_2^N\in\mathbb{N}$ with N large enough such that $[(k_1^N+1)\tau,(k_2^N-1)\tau]\subset(a,b)\subset[k_1^N\tau,k_2^N\tau]$. By squaring (5.6) and summing it from $k=k_1^N+1$ to $k=k_2^N$, there holds

$$\int_a^b \mathcal{J}_{\Lambda,\Omega}(\mu_t^N)\,\mathrm{d}t\le\sum_{k=k_1^N+1}^{k_2^N}\int_{(k-1)\tau}^{k\tau}\mathcal{J}_{\Lambda,\Omega}(\mu_t^{k,N})\,\mathrm{d}t\le\int_a^b|\mathsf{G}_t'|^2\,\mathrm{d}t+2\tau\mathrm{Lip}(\mathsf{G}_t)^2\,. \tag{5.7}$$

By taking $a=0$, $b=1$ in (5.7), we observe that $\int_0^1\mathcal{J}_{\Lambda,\Omega}(\mu_t^N)\,\mathrm{d}t$ is uniformly bounded in N. By Proposition 3.18, up to a subsequence, $\{\mu_t^N\}_{t\in[0,1]}$ weak* converges to a measure $\widetilde{\mu}=(\widetilde{\mathsf{G}},\widetilde{\mathsf{q}},\widetilde{\mathsf{R}})\in\mathcal{CE}_\infty([0,1],\mathsf{G}_0,\mathsf{G}_1)$. Moreover, it follows from (3.38) and (5.7) that, for $[a,b]\subset[0,1]$,

$$\int_a^b\mathcal{J}_{\Lambda,\Omega}(\widetilde{\mu}_t)\,\mathrm{d}t\le\liminf_{N\to+\infty}\int_a^b\mathcal{J}_{\Lambda,\Omega}(\mu_t^N)\,\mathrm{d}t\le\int_a^b|\mathsf{G}_t'|^2\,\mathrm{d}t\,. \tag{5.8}$$

We now show $\widetilde{\mathsf{G}}_t=\mathsf{G}_t$ for $0\le t\le 1$. Note that for any $t\in[0,1]$, there exists a sequence of integers k_N such that $s_N=k_N2^{-N}\to t$ as $N\to\infty$, which implies that $\mathsf{G}_{s_N}^N=\mathsf{G}_{s_N}$ weak* converges to $\widetilde{\mathsf{G}}_t$ by Proposition 3.18. Meanwhile, G_{s_N} weak* converges to G_t by the continuity of G_t. We hence have $\widetilde{\mathsf{G}}_t=\mathsf{G}_t$. Then, it follows from (5.8) that

$$\mathcal{J}_{\Lambda,\Omega}(\widetilde{\mu}_t)=\mathcal{J}_{\Lambda,\Omega}(\mathsf{G}_t,\widetilde{\mathsf{q}}_t,\widetilde{\mathsf{R}}_t)\le|\mathsf{G}_t'|^2\,,$$

by Lebesgue differentiation theorem. The proof of the *only if* direction is completed by noting that (5.4) and (5.5) are invariant with respect to the parameterisation. The uniqueness of $(\mathsf{q}_*,\mathsf{R}_*)$ follows from the linearity of the continuity equation in the variable (q,R) and the strict convexity of the L^2_{G}-norm.

We finally show that when G_t is absolutely continuous with finite 2-energy, $\mu:=(\mathsf{G},\mathsf{G}u_*,\mathsf{G}W_*)\in\mathcal{CE}_\infty([0,1];\mathsf{G}_0,\mathsf{G}_1)$ satisfies $\mathcal{J}_{\Lambda,\Omega}(\mu_t)^{1/2}\le|\mathsf{G}_t'|$ for a.e. $t\in[0,1]$, where (u_*,W_*) is given in Theorem 4.5 (i.e., the Riesz representation of $\widetilde{l}_{\mathsf{G}}$ in $H_{\mathsf{G},\Lambda}(\mathsf{D}^*)$). Let $(a,b)\subset[0,1]$, and $\eta\in C_c^\infty((a,b))$ with $0\le\eta\le1$, and $\{(\mathsf{D}^*\Phi_n\Lambda_1^2,\Phi_n\Lambda_2^2)\}$ with $\Phi_n\in C^1(Q,\mathbb{S}^n)$ be a sequence approximating (u_*,W_*). Then, by using (4.18) and noting $\mathsf{D}^*(\eta^2\Phi)=\eta^2\mathsf{D}^*(\Phi)$, we have

$$\|(\eta u_*,\eta W_*)\|^2_{L^2_{\mathsf{G},\Lambda}(Q)}=\lim_{n\to+\infty}\left\langle(\eta^2u_*,\eta^2W_*),(\mathsf{D}^*\Phi_n\Lambda_1^2,\Phi_n\Lambda_2^2)\right\rangle_{L^2_{\mathsf{G},\Lambda}(Q)}=\lim_{n\to+\infty}l_{\mathsf{G}}(\eta^2\Phi_n)\,. \tag{5.9}$$

By *only if* part proved above, there exists some (q,R) such that

$$\begin{aligned}\left|l_{\mathsf{G}}(\eta^2\Phi_n)\right|&\le\left\|(G^\dagger q,G^\dagger R)\right\|_{L^2_{\mathsf{G},\Lambda}(Q_a^b)}\left\|(\mathsf{D}^*\eta^2\Phi_n,\eta^2\Phi_n)\right\|_{L^2_{\mathsf{G},\Lambda}(Q_a^b)}\\&\le\left(\int_a^b|\mathsf{G}_t'|^2\,\mathrm{d}t\right)^{1/2}\|(\mathsf{D}^*\Phi_n,\Phi_n)\|_{L^2_{\mathsf{G},\Lambda}(Q_a^b)}\,.\end{aligned} \tag{5.10}$$

Combining (5.9) with (5.10) and letting η approximate $\chi_{[a,b]}$, we obtain

$$\|(u_*,W_*)\|_{L^2_{\mathsf{G},\Lambda}(Q_a^b)}\le\left(\int_a^b|\mathsf{G}_t'|^2\,\mathrm{d}t\right)^{1/2}. \tag{5.11}$$

Then, by Lebesgue differentiation theorem again, the inequality (5.11) gives the desired $\mathcal{J}_{\Lambda,\Omega}(\mu_t)^{1/2}\le|\mathsf{G}_t'|$ for the measure $\mu=(\mathsf{G},\mathsf{G}u_*,\mathsf{G}W_*)$. The proof is complete. □

From Lemma 5.1 and Theorem 5.5, we have

$$\begin{aligned}\mathrm{WB}_\Lambda(\mathsf{G}_0,\mathsf{G}_1)&=\inf_{\mathsf{G}}\inf_{(\mathsf{q},\mathsf{R})}\left\{\int_0^1\mathcal{J}_{\Lambda,\Omega}(\mu_t)^{1/2}\,\mathrm{d}t\,;\,\mu=(\mathsf{G},\mathsf{q},\mathsf{R})\in\widetilde{\mathcal{CE}}([0,1];\mathsf{G}_0,\mathsf{G}_1)\right\}\\&=\inf_{\mathsf{G}}\left\{\int_0^1|\mathsf{G}_t'|\,\mathrm{d}t\,;\,\{\mathsf{G}\}_{t\in[0,1]}\text{ is absolutely continuous with }\mathsf{G}_t|_{t=0}=\mathsf{G}_0\,,\,\mathsf{G}_t|_{t=1}=\mathsf{G}_1\right\}.\end{aligned} \tag{5.12}$$

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

Note that if $\{\mu_t\}_{t\in[0,1]}\in\mathcal{CE}_\infty([0,1];\mathsf{G}_0,\mathsf{G}_1)$ minimises $(\mathcal{P})$, then for any $0\le a<b\le 1$, $\{\mu_t\}_{t\in[a,b]}$ is a minimiser to $(\mathcal{P}')$ with $\mathsf{G}_0=\mathsf{G}_t|_{t=a}$ and $\mathsf{G}_1=\mathsf{G}_t|_{t=b}$. Recalling the constant-speed property (5.2) of the minimiser $\mu=(\mathsf{G},\mathsf{q},\mathsf{R})$, we readily see that the associated $\{\mathsf{G}_t\}_{t\in[0,1]}$ is the desired constant-speed geodesic:

$$\mathrm{WB}_\Lambda(\mathsf{G}_s,\mathsf{G}_t)=|t-s|\mathrm{WB}_\Lambda(\mathsf{G}_0,\mathsf{G}_1)\,,\quad \forall 0\le s\le t\le 1\,. \tag{5.13}$$

It allows us to conclude that the inf in (5.12) is attained, and the main result follows.

Corollary 5.7. $(\mathcal{M}(\Omega,\mathbb{S}^n_+),\mathrm{WB}_\Lambda)$ *is a geodesic space. The constant-speed geodesic connecting* $\mathsf{G}_0,\mathsf{G}_1\in\mathcal{M}(\Omega,\mathbb{S}^n_+)$ *is given by the minimiser to* $(\mathcal{P})$.

Another important application of Theorem 5.5 is that we can view the set of $\mathbb{S}^n_+$-valued measures as a pseudo-Riemannian manifold, following [1, Proposition 8.4.5]. We define the tangent space at each $\mathsf{G}\in\mathcal{M}(\Omega,\mathbb{S}^n_+)$ by

$$\begin{aligned}\mathrm{Tan}(\mathsf{G}):=\Big\{&(\mathsf{q},\mathsf{R})\in\mathcal{M}(\Omega,\mathbb{R}^{n\times k}\times\mathbb{M}^n)\,;\,\mathcal{J}_{\Lambda,\Omega}(\mu)<\infty\ \text{ with }\ \mu=(\mathsf{G},\mathsf{q},\mathsf{R})\in\mathcal{M}(\Omega,\mathbb{X});\,.\\ &\mathcal{J}_{\Lambda,\Omega}(\mu)\le\mathcal{J}_{\Lambda,\Omega}((\mathsf{G},\mathsf{q}+\widehat{\mathsf{q}},\mathsf{R}+\widehat{\mathsf{R}}))\,,\ \forall(\widehat{\mathsf{q}},\widehat{\mathsf{R}})\text{ satisfying }\mathsf{D}\widehat{\mathsf{q}}=\widehat{\mathsf{R}}^{\mathrm{sym}}\Big\}\end{aligned} \tag{5.14}$$

From Theorem 5.5, we have that among all the measures (q,R) generating $\{\mathsf{G}_t\}_{t\in[0,1]}$ by the continuity equation, there is a unique one $(\mathsf{q}_*,\mathsf{R}_*)$ with minimal $\mathcal{J}_{\Lambda,\Omega}(\mu_t)$ given by $|\mathsf{G}_t'|$ for a.e. $t\in[0,1]$, that is, $(\mathsf{q}_{*,t},\mathsf{R}_{*,t})\in Tan(\mathsf{G}_t)$ a.e. by (5.14). We also introduce the space $Tan_{field}(\mathsf{G})$ similar to $H_{\mathsf{G},\Lambda}(\mathsf{D}^*)$ (4.20):

$$Tan_{field}(\mathsf{G})=\overline{\left\{(\mathsf{D}^*\Phi\Lambda_1^2,\,\Phi\Lambda_2^2)\,;\,\Phi\in C^1(\Omega,\mathbb{S}^n)\right\}}^{\|\cdot\|_{L^2_{\mathsf{G},\Lambda}(\Omega)}}\,.$$

Then, similarly to the argument for Theorem 4.5, the tangent space $Tan(\mathsf{G})$ can be characterised as follows:

$$(\mathsf{q},\mathsf{R})\in Tan(\mathsf{G})\quad\text{if and only if}\quad(\mathsf{q},\mathsf{R})=\mathsf{G}(u,W)\text{ with }(u,W)\in Tan_{field}(\mathsf{G})\,. \tag{5.15}$$

We summarise the above discussions in the following corollary, which provides a Riemannian interpretation of the transport distance $\mathrm{WB}_\Lambda(\cdot,\cdot)$.

Corollary 5.8. *Let* $\{\mathsf{G}_t\}_{t\in[0,1]}$ *be an absolutely continuous curve in* $(\mathcal{M}(\Omega,\mathbb{S}^n_+),\mathrm{WB}_\Lambda)$ *and* $\{(\mathsf{q}_t,\mathsf{R}_t)\}_{t\in[0,1]}$ *be the family of measures in* $\mathcal{M}(\Omega,\mathbb{R}^{n\times k}\times\mathbb{M}^n)$ *such that* $\mu=(\mathsf{G},\mathsf{q},\mathsf{R})\in\mathcal{CE}([0,1];\mathsf{G}_0,\mathsf{G}_1)$ *and* $\mathcal{J}_{\Lambda,\Omega}(\mu_t)$ *is finite a.e.. Then* $|\mathsf{G}_t'|=\mathcal{J}_{\Lambda,\Omega}(\mu_t)$ *holds for a.e.* $t\in[0,1]$ *if and only if* $(\mathsf{q}_t,\mathsf{R}_t)\in Tan(\mathsf{G}_t)$ *a.e., where* $Tan(\mathsf{G})$ *is defined in (5.14) and characterised by (5.15). Moreover, for absolutely continuous* G_t *with finite 2-energy (i.e.,* $\mathsf{G}\in\mathcal{AC}([0,1];\mathsf{G}_0,\mathsf{G}_1)$*), let* (u_*,W_*) *be the unique minimiser to (4.22). Then, there holds* $(u_{*,t},W_{*,t})\in Tan_{field}(\mathsf{G}_t)$ *a.e..*

6. Cone space and spherical distance

In this section, we discuss the conic structure of our weighted transport distance WB_Λ, which extends the results in [16, Section 4] and [73, Section 5]. The starting point is a spherical distance associated with WB_Λ:

$$\mathrm{SWB}^2_\Lambda(\mathsf{G}_0,\mathsf{G}_1)=\inf\left\{\mathcal{J}_{\Lambda,Q}(\mu)\,;\,\mu\in\widetilde{\mathcal{CE}}([0,1];\mathsf{G}_0,\mathsf{G}_1)\,,Tr_\Lambda\mathsf{G}_t(\Omega)=1\right\},\quad\text{for }\mathsf{G}_0,\mathsf{G}_1\in\mathcal{M}_1\,, \tag{6.1}$$

where $Tr_\Lambda(X):=Tr\big(\widetilde{\Lambda}_2^{-1}X\widetilde{\Lambda}_2^{-1}\big)$ with $\widetilde{\Lambda}_2=n\Lambda_2/Tr(\Lambda_2)$ is the scaled trace and

$$\mathcal{M}_1:=\{\mathsf{G}\in\mathcal{M}(\Omega,\mathbb{S}^n_+)\,;\,Tr_\Lambda\mathsf{G}(\Omega)=1\}\,. \tag{6.2}$$

We will prove that $(\mathcal{M}_1,\mathrm{SWB}_\Lambda)$ is a complete geodesic space and $(\mathcal{M}(\Omega,\mathbb{S}^n_+),\mathrm{WB}_\Lambda)$ can be viewed as its metric cone. Let us first recall some basic concepts [19, 60]. We consider a metric space (X,d_X) with diameter $\mathrm{diam}(X)=\sup_{x,y\in X}d_X(x,y)\le\pi$. The associated cone is defined by $\mathfrak{C}(X):=X\times[0,\infty)\backslash X\times\{0\}$ with the metric

$$d^2_{\mathfrak{C}(X)}([x_0,r_0],[x_1,r_1]):=r_0^2+r_1^2-2r_0r_1\cos\left(d_X(x_0,x_1)\right), \tag{6.3}$$

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

where a point in $\mathfrak{C}(X)$ is of the form $[x, r]$ with $x \in X$ and $r \geq 0$ and satisfies the equivalence relation $[x_0, 0] \sim [x_1, 0]$. It can be proved that for $x_0, x_1 \in X$ with $0 < d_X(x_0, x_1) < \pi$ and $r_0, r_1 > 0$, there is one-to-one correspondence between the geodesics for $d_{\mathfrak{C}(X)}([x_0, r_0], [x_1, r_1])$ and for $d_X(x_0, x_1)$; see [60, Theorem 2.6]. In particular, we have the following useful lemmas from [16, Lemma 4.4] and [60, Theorem 2.2], respectively.

Lemma 6.1. *If X is a length space, then the distance $d_X(x_0, x_1)$ can be characterised by*

$$d_X(x_0, x_1) = \inf\left\{ \int_0^1 \left|[x_t, 1]'\right|_{\mathfrak{C}(X)} \mathrm{d}t\,;\ [x_t, 1] \text{ is absolutely continuous and connects } [x_0, 1] \text{ and } [x_1, 1]\right\},$$

where $|[x_t, 1]'|_{\mathfrak{C}(X)}$ is the metric derivative in the space $(\mathfrak{C}(X), d_{\mathfrak{C}(X)})$.

Lemma 6.2. *Let $\mathfrak{C}(X)$ be the cone as above and $(\mathfrak{C}(X), d)$ be a metric space for some metric d. If there holds*

$$d^2([x_0, r_0], [x_1, r_1]) = r_0 r_1 d^2([x_0, 1], [x_1, 1]) + (r_0 - r_1)^2\,, \tag{6.4}$$

and $0 < d^2([x_0, 1], [x_1, 1]) \leq 4$ for $x_0 \neq x_1$, then $d_X(x_0, x_1) := \arccos\left(1 - d^2([x_0, 1], [x_1, 1])/2\right)$ is a metric on X such that (6.3) holds, equivalently, $(\mathfrak{C}(X), d)$ is a metric cone over (X, d_X).

We are now ready to consider the conic properties of $(\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda)$. For this, we set $r := \sqrt{Tr_\Lambda(\mathsf{G}(\Omega))} \geq 0$ for a measure $\mathsf{G} \in \mathcal{M}(\Omega, \mathbb{S}^n_+)$ and identify G with $[\mathsf{G}/r^2, r] \in \mathfrak{C}(\mathcal{M}_1)$.

Theorem 6.3. *Suppose that there holds $\mathsf{D}^*(\Lambda_2^{-2}) = 0$ and let $c := \sqrt{2}n/Tr(\Lambda_2)$. Then, $(\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda/c)$ is a metric cone over $(\mathcal{M}_1, \mathrm{SWB}_\Lambda/c)$, namely, for $\mathsf{G}_0, \mathsf{G}_1 \in \mathcal{M}_1$ and $r_0, r_1 \geq 0$,*

$$\mathrm{WB}^2_\Lambda(r_0^2 \mathsf{G}_0, r_1^2 \mathsf{G}_1)/c^2 = r_0^2 + r_1^2 - 2 r_0 r_1 \cos\left(\mathrm{SWB}_\Lambda(\mathsf{G}_0, \mathsf{G}_1)/c\right), \tag{6.5}$$

and $(\mathcal{M}_1, \mathrm{SWB}_\Lambda/c)$ is a complete geodesic space with $\mathrm{diam}(\mathcal{M}_1) \leq \pi$.

Proof. We first prove that $(\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda/c)$ is a metric cone over $(\mathcal{M}_1, d)$ for some metric d. For this, we note from (3.18) in the proof of Lemma 3.9 that

$$\mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1) \leq 2 \int_\Omega \left\| \left(\sqrt{G_1} - \sqrt{G_0}\right) \Lambda_2^{-1} \right\|_{\mathrm{F}}^2 \mathrm{d}\lambda \leq 4\big(n/Tr(\Lambda_2)\big)^2 \big(Tr_\Lambda \mathsf{G}_0(\Omega) + Tr_\Lambda \mathsf{G}_1(\Omega)\big)\,,$$

which yields $\mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1) \leq 4c^2$ for $\mathsf{G}_0, \mathsf{G}_1 \in \mathcal{M}_1$. By Lemma 6.2, it suffices to check the scaling property (6.4):

$$\mathrm{WB}^2_\Lambda(r_0^2 \mathsf{G}_0, r_1^2 \mathsf{G}_1)/c^2 = r_0 r_1 \mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1)/c^2 + (r_0 - r_1)^2\,, \tag{6.6}$$

for $\mathsf{G}_0, \mathsf{G}_1 \in \mathcal{M}_1$ and $r_0, r_1 \geq 0$ to show that $(\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda/c)$ is a metric cone. Note that (6.6) for the case of $r_0 = 0$ or $r_1 = 0$ follows from Proposition 4.4. Thus, we can assume $r_0, r_1 > 0$. Let $\{\mu_t = (\mathsf{G}_t, \mathsf{q}_t, \mathsf{R}_t)\}_{t \in [0,1]} \in \widetilde{\mathcal{CE}}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$ be an admissible curve. We define scalar functions $b(t) = r_0 + (r_1 - r_0)t$ and $a(t) := t r_1 / b(t)$. It is clear that $a(t)$ is strictly increasing with inverse denoted by $t(a)$. We then define $\widetilde{\mathsf{G}}_t = b(t)^2 \mathsf{G}_{a(t)}$ with

$$\widetilde{\mathsf{q}}_t = a'(t) b(t)^2 \mathsf{q}_{a(t)}\,, \quad \widetilde{\mathsf{R}}_t = a'(t) b(t)^2 \mathsf{R}_{a(t)} + 2b(t)(r_1 - r_0)\mathsf{G}_{a(t)}\,,$$

which satisfies the continuity equation with end points $r_0^2 \mathsf{G}_0$ and $r_1^2 \mathsf{G}_1$. We now compute

$$\begin{aligned} \mathcal{J}_{\Lambda,\Omega}\big(\widetilde{\mathsf{G}}, \widetilde{\mathsf{q}}, \widetilde{\mathsf{R}}\big) = & \int_0^1 a'(t(a))\, b(t(a))^2 \mathcal{J}_{\Lambda,\Omega}(\mathsf{G}_a, \mathsf{q}_a, \mathsf{R}_a)\, \mathrm{d}a + c^2 (r_1 - r_0)^2 \int_0^1 Tr_\Lambda \mathsf{G}_{a(t)}(\Omega)\, \mathrm{d}t \\ & + c^2 \int_0^1 b(t(a))\, (r_1 - r_0) Tr_\Lambda \mathsf{R}_a(\Omega)\, \mathrm{d}a\,. \end{aligned} \tag{6.7}$$

The last two terms in (6.7) can be simplified by (3.13) on $[0, 1]$ with test function $\Phi_s = b(t(s))\, \Lambda_2^{-2}$:

$$\int_0^1 t'(a)(r_1 - r_0) Tr_\Lambda \mathsf{G}_a(\Omega) + b(t(a)) Tr_\Lambda \mathsf{R}_a(\Omega)\, \mathrm{d}a = r_1 Tr_\Lambda \mathsf{G}_1(\Omega) - r_0 Tr_\Lambda \mathsf{G}_0(\Omega)\,,$$

which implies, thanks to $Tr_\Lambda \mathsf{G}_0(\Omega) = Tr_\Lambda \mathsf{G}_1(\Omega) = 1$,

$$\int_0^1 (r_1 - r_0)^2 Tr_\Lambda \mathsf{G}_{a(t)}(\Omega)\,\mathrm{d}t + \int_0^1 b(t(a))(r_1 - r_0) Tr_\Lambda \mathsf{R}_a(\Omega)\,\mathrm{d}a = (r_1 - r_0)^2 . \tag{6.8}$$

Therefore, by noting $a'(t)b(t)^2 = r_0 r_1$ and using (6.8), it follows that

$$\mathcal{J}_{\Lambda,\mathcal{Q}}(\widetilde{\mathsf{G}}, \widetilde{\mathsf{q}}, \widetilde{\mathsf{R}}) = r_0 r_1 \int_0^1 \mathcal{J}_{\Lambda,\Omega}(\mathsf{G}_a, \mathsf{q}_a, \mathsf{R}_a)\,\mathrm{d}a + c^2(r_1 - r_0)^2 ,$$

which readily gives $\mathrm{WB}^2_\Lambda(r_0^2\mathsf{G}_0, r_1^2\mathsf{G}_1)/c^2 \le r_0 r_1 \mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1)/c^2 + (r_0 - r_1)^2$. The other direction can be proved similarly. We have proved the existence of $(\mathcal{M}_1, d)$ such that $(\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda/c)$ is the associated metric cone.

We now show that the metric d on $\mathcal{M}_1$ is given by SWB_Λ/c.

By Corollary 5.7 and [18, Corollary 5.11], we have that $(\mathcal{M}_1, d)$ is a geodesic space, which, by Lemma 6.1, gives, for $\mathsf{G}_0, \mathsf{G}_1 \in \mathcal{M}_1$,

$$d(\mathsf{G}_0, \mathsf{G}_1) = \inf\left\{ \int_0^1 |\mathsf{G}_t'|\,\mathrm{d}t\,;\ \mathsf{G}_t \text{ is absolutely continuous in } (\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda/c) \text{ with } \mathsf{G}_t \in \mathcal{M}_1 \right\}.$$

It then follows from Theorem 5.5 and definition (6.1) that $d(\mathsf{G}_0, \mathsf{G}_1) = \mathrm{SWB}_\Lambda(\mathsf{G}_0, \mathsf{G}_1)/c$ and hence (6.5) holds. Recalling $\mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1)/c^2 \le 4$ for $\mathsf{G}_0, \mathsf{G}_1 \in \mathcal{M}_1$, (6.5) gives $0 \le \mathrm{SWB}_\Lambda(\mathsf{G}_0, \mathsf{G}_1)/c \le \pi$. Finally, for the completeness of $(\mathcal{M}_1, \mathrm{SWB}_\Lambda/c)$, it suffices to note that SWB_Λ and WB_Λ are topologically equivalent on $\mathcal{M}_1$, again by (6.5), and $\mathcal{M}_1$ is a closed set in $(\mathcal{M}(\Omega, \mathbb{S}^n_+), \mathrm{WB}_\Lambda)$ by Proposition 5.2. □

7. Example and discussion

In this section, we detail the connections between our model ($\mathcal{P}$) and the existing ones.

Example 7.1. *(Kantorovich–Bures metric [16]). We set the dimension parameters $n = m = d$ and $k = 1$ and the weight matrices $\Lambda_i = I$ for $i = 1, 2$ in (3.1) and consider the differential operator $\mathsf{D} = \nabla_s$ for the continuity equation (3.13), where ∇_s is the symmetric gradient defined by $\nabla_s(q) = \frac{1}{2}(\nabla q + (\nabla q)^{\mathrm{T}})$ for a smooth vector field $q \in C_c^\infty(\mathbb{R}^d, \mathbb{R}^d)$. Then, ($\mathcal{P}$) gives the convex formulation of the Kantorovich–Bures metric d_{KB} on $\mathcal{M}(\Omega, \mathbb{S}^d_+)$ [16, Definition 2.1]:*

$$\begin{aligned}\mathrm{WB}^2_{(I,I)}(\mathsf{G}_0, \mathsf{G}_1) = \frac{1}{2} d^2_{KB}(\mathsf{G}_0, \mathsf{G}_1) = \inf\big\{ &\mathcal{J}_{\Lambda,\mathcal{Q}}(\mu)\,;\ \mu = (\mathsf{G},\mathsf{q},\mathsf{R}) \in \mathcal{M}(\mathcal{Q}, \mathbb{X}) \text{ satisfies} \\ &\partial_t \mathsf{G} = \{-\nabla \mathsf{q}_t + \mathsf{R}_t\}^{\mathrm{sym}} \text{ with } \mathsf{G}_t|_{t=0} = \mathsf{G}_0\,,\ \mathsf{G}_t|_{t=1} = \mathsf{G}_1 \big\}, \end{aligned} \tag{$\mathcal{P}_{\mathrm{WB}}$}$$

for $\mathsf{G}_0, \mathsf{G}_1 \in \mathcal{M}(\Omega, \mathbb{S}^d_+)$, where $\mathcal{J}_{\Lambda,\mathcal{Q}}(\mu)$ with $\Lambda = (I, I)$ is given by (3.24):

$$\mathcal{J}_{\Lambda,\mathcal{Q}}(\mu) = \frac{1}{2}\|G^\dagger q\|^2_{L^2_{\mathsf{G}}(\mathcal{Q})} + \frac{1}{2}\|G^\dagger R\|^2_{L^2_{\mathsf{G}}(\mathcal{Q})} .$$

Example 7.2. *(Wasserstein–Fisher–Rao metric [27, 56, 64]). If we set $n = m = 1$, $k = d$ and $\Lambda_1 = \sqrt{\alpha} I$, $\Lambda_2 = \sqrt{\beta} I$ with $\alpha, \beta > 0$, and consider the differential operator $\mathsf{D} = \mathrm{div}$, then ($\mathcal{P}$) gives the Wasserstein–Fisher–Rao metric [64, (3.1)]: for given distributions $\rho_0, \rho_1 \in \mathcal{M}(\Omega, \mathbb{R}_+)$,*

$$\mathrm{WFR}^2(\rho_0, \rho_1) = \inf\left\{ \int_0^1 \int_\Omega \rho^\dagger\left(\frac{1}{2\alpha}|q|^2 + \frac{1}{2\beta} r^2\right)\mathrm{d}x\,\mathrm{d}t\,;\ \partial_t\rho + \mathrm{div}\, q = r \text{ with } \rho_t|_{t=0} = \rho_0\,,\ \rho_t|_{t=1} = \rho_1 \right\}. \tag{$\mathcal{P}_{\mathrm{WFR}}$}$$

Example 7.3. *(Matricial interpolation distance [25]). Let N be a positive integer and $(\mathbb{M}^n)^N$ denote the space of block-row vectors $(A_1, \ldots, A_N)$ with $A_i \in \mathbb{M}^n$. The spaces $(\mathbb{S}^n)^N$ and $(\mathbb{A}^n)^N$ are defined similarly. For $M \in (\mathbb{M}^n)^N$, we define its component transpose by $M^t := (M_1^{\mathrm{T}}, \ldots, M_N^{\mathrm{T}})$. We fix a sequence of symmetric matrices $\{L_k\}_{k=1}^N \subset \mathbb{S}^n$ and define the linear operator $\nabla_L : \mathbb{S}^n \to (\mathbb{A}^n)^N$ by $(\nabla_L X)_k = L_k X - X L_k$. We* *denote by* ∇_L^* *its dual operator with respect to the Frobenius inner product. We now let* $k = n(d+N)$ *and write* $\mathsf{q} \in \mathcal{M}(Q, \mathbb{R}^{n\times k})$ *for* $[\mathsf{q}_0, \mathsf{q}_1]$ *with* $\mathsf{q}_0 \in \mathcal{M}(Q, (\mathbb{M}^n)^d)$ *and* $\mathsf{q}_1 \in \mathcal{M}(Q, (\mathbb{M}^n)^N)$*. With the above notions, we define*

$$\mathsf{D}\,\mathsf{q} := \frac{1}{2}\mathrm{div}(\mathsf{q}_0 + \mathsf{q}_0^t) - \frac{1}{2}\nabla_L^*(\mathsf{q}_1 - \mathsf{q}_1^t)\,.$$

Then, it is clear that ($\mathcal{P}$) *with weight matrices* $\Lambda_i = I$ *for* $i = 1, 2$ *gives the model in [25, (5.7a)–(5.7c)]:*

$$\mathrm{W}_{2,\mathrm{FR}}(\mathsf{G}_0, \mathsf{G}_1)^2 = \frac{1}{2}\inf\Big\{\|G^\dagger q_0\|^2_{L^2_{\mathsf{G}(Q)}} + \|G^\dagger q_1\|^2_{L^2_{\mathsf{G}(Q)}} + \|G^\dagger R\|^2_{L^2_{\mathsf{G}(Q)}}\,;$$
$$\partial_t \mathsf{G} = -\frac{1}{2}\mathrm{div}(\mathsf{q}_0 + \mathsf{q}_0^t) + \frac{1}{2}\nabla_L^*(\mathsf{q}_1 - \mathsf{q}_1^t) + \mathsf{R}^{\mathrm{sym}} \textit{ with } \mathsf{G}_t|_{t=0} = \mathsf{G}_0\,,\ \mathsf{G}_t|_{t=1} = \mathsf{G}_1\Big\}\,. \qquad (\mathcal{P}_{2,\mathrm{FR}})$$

We next relate our model ($\mathcal{P}$) to the matrix-valued optimal ballistic transport problems in refs. [15, 91]. As reviewed in the introduction, Brenier [15] recently attempted to find the weak solution of the incompressible Euler equation on the domain $[0, T] \times \Omega \subset \mathbb{R}^{1+d}$ (we omit the initial and boundary conditions for simplicity):

$$\partial_t v + \mathrm{div}\,(v \otimes v) + \nabla p = 0\,, \quad \mathrm{div}\, v = 0\,, \qquad (7.1)$$

by minimising the kinetic energy $\int_0^T \int_\Omega |v(t,x)|^2\,\mathrm{d}x\,\mathrm{d}t$, where v is a $\mathbb{R}^n$-valued vector field and p is a scalar function. It turns out that this problem admits a concave maximisation dual problem, to which the relaxed solution always exists under very light assumptions. Such an approach was extended by Vorotnikov [91] in an abstract functional analytic framework that includes a broad class of PDEs with quadratic nonlinearity as examples, such as the Hamilton–Jacobi equation, the template matching equation, and the multidimensional Camassa–Holm equation. More precisely, [91] considered the following abstract Euler equation on $[0, T] \times \Omega$:

$$\partial_t v = \mathsf{P} \circ \mathsf{L}\,(v \otimes v)\,, \quad v(0,\cdot) = v_0 \in \mathsf{P}(L^2(\Omega, \mathbb{R}^n))\,, \qquad (7.2)$$

where P is an orthogonal projection and $\mathsf{L}: L^2(\Omega, \mathbb{S}^n) \to L^2(\Omega, \mathbb{R}^n)$ is a (closed densely defined) linear operator. One can see that for $\mathsf{L} = -\mathrm{div}$ and P being the Leray projection, the problem (7.2) reduces to (7.1). The dual problem associated with the weak solution of (7.2) with minimal kinetic energy reads as follows:

$$\sup\left\{\int_0^T \int_\Omega v_0 \cdot q - \frac{1}{2} q \cdot G^\dagger q\,\mathrm{d}x\,\mathrm{d}t\,;\ \partial_t G + 2(\mathsf{L}^* \circ \mathsf{P})\,q = 0 \text{ with } G(T) = I\right\}, \qquad (7.3)$$

where G and q are $\mathbb{S}^n_+$-valued and $\mathbb{R}^n$-valued vector fields, respectively. Note that the Hamilton–Jacobi equation $\partial_t \psi + \frac{1}{2}|\nabla\psi|^2 = 0$ can be reformulated as $\partial_t v + \frac{1}{2}\nabla Tr(v \otimes v) = 0$ by letting $v = \nabla\psi$, which is a special case of (7.2) with $\mathsf{P} = I$ and $\mathsf{L} = -\frac{1}{2}\nabla Tr$. The corresponding dual maximisation problem is given by

$$\sup\left\{-\int_\Omega \psi_0\rho_0\,\mathrm{d}x - \frac{1}{2}\int_0^T \int_\Omega \rho^\dagger |q|^2\,\mathrm{d}x\,\mathrm{d}t\,;\ \partial_t\rho + \mathrm{div}\,q = 0 \text{ with } \rho(T) = 1\right\}, \qquad (7.4)$$

which closely relates to the ballistic transport problem [5]. In view of (7.3) and (7.4), one may regard

$$\partial_t G + 2(\mathsf{L}^* \circ \mathsf{P})\,q = 0 \qquad (7.5)$$

as a matricial continuity equation, and our model (3.14) can be hence viewed as an unbalanced variant of (7.5). Then, the conservativity condition $\mathsf{D}^*(I) = 0$ for (7.5) is simply $\mathsf{P} \circ \mathsf{L}(I) = 0$, which has been used to guarantee the existence of a measure-valued solution to (7.3); see [91, Theorem 4.6]. Thanks to the above observations, one may expect that each meaningful choice of L and P in [91, Section 6] can generate a reasonable distance ($\mathcal{P}$) with $\mathsf{D} = 2(\mathsf{L}^* \circ \mathsf{P})$. For instance, setting $n = d$, $\mathsf{P} = I$ and $\mathsf{L} = -\mathrm{div} - \frac{1}{2}\nabla Tr$ in (7.2) gives the template matching equation $\partial_t v + \mathrm{div}\,(v \otimes v) + \frac{1}{2}\nabla|v|^2 = 0$ and a distance ($\mathcal{P}$) with $\mathsf{D} = 2(\mathsf{L}^* \circ \mathsf{P})$:

$$\inf\left\{\mathcal{J}_{\Lambda,Q}(\mathsf{G},\mathsf{q},\mathsf{R})\,;\ \partial_t \mathsf{G} + 2\nabla_s \mathsf{q} + \mathrm{div}\mathsf{q} I = \mathsf{R}_t^{\mathrm{sym}} \text{ with } \mathsf{G}_t|_{t=0} = \mathsf{G}_0\,,\ \mathsf{G}_t|_{t=1} = \mathsf{G}_1\right\}. \qquad (7.6)$$

Remark 7.1. *An important question is how to compare these matrix-valued OT models ($\mathcal{P}_{\mathrm{WB}}$), ($\mathcal{P}_{2,\mathrm{FR}}$), and (7.6) (as well as others in the literature), which requires a deeper theoretical analysis and is completely open, to the best of our knowledge.*

8. Concluding remarks

We have proposed a general class of unbalanced matrix-valued OT distances $\mathrm{WB}_\Lambda(\cdot,\cdot)$ over the space $\mathcal{M}(\Omega, \mathbb{S}^n_+)$, called the weighted Wasserstein–Bures metric. The definition relies on a dynamic formulation and convex analysis. We have shown that $\mathcal{M}(\Omega, \mathbb{S}^n_+)$ equipped with the metric $\mathrm{WB}_\Lambda(\cdot,\cdot)$ is a complete geodesic space, and it can be viewed as a metric cone. In the follow-up work [63], we have considered the convergence of the discrete approximation of the transport model ($\mathcal{P}$). Our results provide a unified framework for unbalanced transport distances on matrix-valued measures and directly apply to various existing models such as the Kantorovich–Bures distance ($\mathcal{P}_{\mathrm{WB}}$), the matricial interpolation distance ($\mathcal{P}_{2,\mathrm{FR}}$) and the WFR one ($\mathcal{P}_{\mathrm{WFR}}$). Meanwhile, it paves the way for practical applications, in particular, diffusion tensor imaging as in refs. [26, 77, 86].

Acknowledgements. The authors would like to thank the anonymous referees and editors for their careful reading and constructive comments and suggestions, which have helped us improve this work.

Financial Support. The work of Bowen Li is supported in part by National Key R&D Program of China (project 2024YFA1016000). Jun Zou was substantially supported by Hong Kong RGC General Research Fund (projects 14308322 and 14306921) and NSFC/Hong Kong RGC Joint Research Scheme 2022/23 (project N_CUHK465/22).

Competing interests. The authors declare none.

References

[1] Ambrosio, L., Gigli, N. & Savaré, G. (2008) *Gradient Flows: In Metric Spaces and in the Space of Probability Measures. Lectures in Mathematics ETH Zürich*, 2nd ed., Basel, Birkhäuser Verlag.
[2] Ambrosio, L. & Tilli, P. (2004) *Topics on Analysis in Metric Spaces. Oxford Lecture Series in Mathematics and its Applications*, Vol. **25**, Oxford, Oxford University Press.
[3] Arjovsky, M., Chintala, S. & Bottou, L. (2017). Wasserstein generative adversarial networks. In International Conference on Machine Learning, PMLR, pp. 214–223.
[4] Barbu, V. & Precupanu, T. (2012) *Convexity and Optimization in Banach Spaces. Springer Monographs in Mathematics*, 4th ed., Dordrecht, Springer.
[5] Barton, A. & Ghoussoub, N. (2019) Dynamic and stochastic propagation of the brenier optimal mass transport. *Eur. J. Appl. Math.* **30**(6), 1264–1299.
[6] Bauschke, H. H. & Combettes, P. L. (2011) *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Vol. **408**, Berlin, Springer.
[7] Benamou, J.-D. (2003) Numerical resolution of an “unbalanced” mass transport problem. *ESAIM: Math. Modell. Numer. Anal.* **37**(5), 851–868.
[8] Benamou, J.-D. & Brenier, Y. (2000) A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numer. Math.* **84**(3), 375–393.
[9] Bhatia, R. (1997) *Matrix Analysis, Grad. Texts in Math*, Vol. **169**, New York, Springer-Verlag.
[10] Bhatia, R., Jain, T. & Lim, Y. (2019) On the bures–wasserstein distance between positive definite matrices. *Expo. Math.* **37**(2), 165–191.
[11] Bonafini, M., Orlandi, G. & Oudet, É. (2018) Variational approximation of functionals defined on 1-dimensional connected sets: The planar case. *SIAM J. Math. Anal.* **50**(6), 6307–6332.
[12] Bouchitté, G. (2020) Convex analysis and duality, arXiv preprint arXiv: 2004.09330.
[13] Bouchitté, G. & Valadier, M. (1988) Integral representation of convex functionals on a space of measures. *J. Funct. Anal.* **80**(2), 398–420.
[14] Brenier, Y. (1991) Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.* **44**(4), 375–417.
[15] Brenier, Y. (2018) The initial value problem for the euler equations of incompressible fluids viewed as a concave maximization problem. *Commun. Math. Phys.* **364**(2), 579–605.
[16] Brenier, Y. & Vorotnikov, D. (2020) On optimal transport of matrix-valued measures. *SIAM J. Math. Anal.* **52**(3), 2849–2873.

[17] Brezis, H. (2010) *Functional Analysis, Sobolev Spaces and Partial Differential Equations, Universitext*, New York, NY, Springer. (English)
[18] Bridson, M. R. & Haefliger, A. (2013) *Metric Spaces of Non-Positive Curvature*, Vol. **319**, Berlin, Germay, Springer.
[19] Burago, D., Burago, Y. & Ivanov, S. (2022) *A Course in Metric Geometry, Volume 33 of Graduate Studies in Mathematics*, Providence, RI, American Mathematical Society.
[20] Caffarelli, L. A. & McCann, R. J. (2010) Free boundaries in optimal transport and monge-ampere obstacle problems. *Ann. Math.* **171**(2), 673–730.
[21] Carlen, E. A. & Maas, J. (2014) An analog of the 2-wasserstein metric in non-commutative probability under which the fermionic fokker–planck equation is gradient flow for the entropy. *Commun. Math. Phys.* **331**(3), 887–926.
[22] Carlen, E. A. & Maas, J. (2017) Gradient flow and entropy inequalities for quantum markov semigroups with detailed balance. *J. Funct. Anal.* **273**(5), 1810–1869.
[23] Chen, Y., Gangbo, W., Georgiou, T. T. & Tannenbaum, A. (2020) On the matrix monge–kantorovich problem. *Eur. J. Appl. Math.* **31**(4), 574–600.
[24] Chen, Y., Georgiou, T. T. & Tannenbaum, A. (2017) Matrix optimal mass transport: A quantum mechanical approach. *IEEE Trans. Automat. Control* **63**(8), 2612–2619.
[25] Chen, Y., Georgiou, T. T. & Tannenbaum, A. (2019) Interpolation of matrices and matrix-valued densities: The unbalanced case. *Eur. J. Appl. Math.* **30**(3), 458–480.
[26] Chen, Y., Haber, E., Yamamoto, K., Georgiou, T. T. & Tannenbaum, A. (2018) An efficient algorithm for matrix-valued and vector-valued optimal mass transport. *J. Sci. Comput.* **77**(1), 79–100.
[27] Chizat, L., Peyré, G., Schmitzer, B. & Vialard, F.-X. (2018) An interpolating distance between optimal transport and fisher–rao metrics. *Found. Comput. Math.* **18**(1), 1–44.
[28] Chizat, L., Peyré, G., Schmitzer, B. & Vialard, F.-X. (2018) Scaling algorithms for unbalanced optimal transport problems. *Math. Comput.* **87**(314), 2563–2609.
[29] Chizat, L., Peyré, G., Schmitzer, B. & Vialard, F.-X. (2018) Unbalanced optimal transport: Dynamic and kantorovich formulations. *J. Funct. Anal.* **274**(11), 3090–3123.
[30] Cole, S., Eckstein, M., Friedland, S. & Życzkowski, K. (2023) On quantum optimal transport. *Math. Phys. Anal. Geomet.* **26**(2), 14.
[31] Datta, N. & Rouzé, C. (2020) Relating relative entropy, optimal transport and fisher information: A quantum hwi inequality. *Ann. Henri Poincaré* **21**(7), 2115–2150.
[32] De Palma, G., Marvian, M., Trevisan, D. & Lloyd, S. (2021) The quantum wasserstein distance of order 1. *IEEE Trans. Inform. Theory* **67**(10), 6627–6643.
[33] De Palma, G. & Trevisan, D. (2021). Quantum optimal transport with quantum channels. In *Annales Henri Poincaré*, Vol. 22, Springer, pp. 3199–3234.
[34] Dolbeault, J., Nazaret, B. & Savaré, G. (2009) A new class of transport distances between measures. *Calc. Var. Partial Differ. Eq.* **34**(2), 193–231.
[35] Duran, A. J. & Lopez-Rodriguez, P. (1997) The lpspace of a positive definite matrix of measures and density of matrix polynomials inl1. *J. Approx. Theory* **90**(2), 299–318.
[36] Evans, L. C. & Gariepy, R. F. (2015) *Measure Theory and Fine Properties of Functions, Textbooks in Mathematics*, Boca Raton, FL, CRC Press.
[37] Ferradans, S., Papadakis, N., Peyré, G. & Aujol, J.-F. (2014) Regularized discrete optimal transport. *SIAM J. Imaging Sci.* **7**(3), 1853–1882.
[38] Figalli, A. (2010) The optimal partial transport problem. *Arch. Ration. Mech. Anal.* **195**(2), 533–560.
[39] Figalli, A. & Gigli, N. (2010) A new transportation distance between non-negative measures, with applications to gradients flows with dirichlet boundary conditions. *J. math. pures appl.* **94**(2), 107–130.
[40] Flaherty, F. & do Carmo, M. (2013). *Riemannian Geometry. Mathematics: Theory & Applications*, Birkhäuser Boston.
[41] Fleissner, F. C. (2021) A minimizing movement approach to a class of scalar reaction–diffusion equations. *ESAIM: Control Optim. Calc. Var.* **27**, 18.
[42] Gerald, B. (1999) *Folland Real Analysis: Modern Techniques and Their Applications (second edition), Pure and Applied Mathematics (New York)*, New York, John Wiley & Sons, Inc.
[43] Frogner, C., Zhang, C., Mobahi, H., Araya, M. & Poggio, T. A. (2015) Learning with a wasserstein loss. *Adv. Neur. Inf. Process. Syst.* **28**, 2053–2061.
[44] Gallouët, T. O. & Monsaingeon, L. (2017) A jko splitting scheme for kantorovich–fisher–rao gradient flows. *SIAM J. Math. Anal.* **49**(2), 1100–1130.
[45] Golse, F., Mouhot, C. & Paul, T. (2016) On the mean field and classical limits of quantum mechanics. *Commun. Math. Phys.* **343**(1), 165–205.
[46] Golse, F. & Paul, T. (2017) The schrödinger equation in the mean-field and semiclassical regime. *Arch. Ration. Mech. Anal.* **223**(1), 57–94.
[47] Golse, F. & Paul, T. (2018) Wave packets and the quadratic monge–kantorovich distance in quantum mechanics. *C. R. Math.* **356**(2), 177–197.
[48] Gross, L. (1975) Hypercontractivity and logarithmic sobolev inequalities for the clifford-dirichlet form. *Duke Math. J.* **42**(3), 383–396.
[49] Guittet, K. (2002) Extended kantorovich norms: A tool for optimization. *Technical Report 4402*.
[50] Hanin, L. G. (1999) An extension of the kantorovich norm. *Contemp. Math.* **226**, 113–130.

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

[51] Jordan, R., Kinderlehrer, D. & Otto, F. (1998) The variational formulation of the fokker–planck equation. *SIAM J. Math. Anal.* **29**(1), 1–17.
[52] Kantorovich, L. V. (1942). On the translocation of masses. In: *Dokl. Aauk. USSR (NS)*, Vol. 37, pp. 199–201.
[53] Kantorovich, L. V. & Rubinshtein, G. S. (1957) On a functional space and certain extremum problems. *Dokl. Akad. Nauk.* **115**(6), 1058–1061.
[54] Kantorovich, L. V. & Rubinshtein, S. (1958) On a space of totally additive functions. *Vestn. St. Petersburg Univ. Math.* **13**(7), 52–59.
[55] Kastoryano, M. J. & Temme, K. (2013) Quantum logarithmic sobolev inequalities and rapid mixing. *J. Math. Phys.* **54**(5), 052202.
[56] Kondratyev, S., Monsaingeon, L., Vorotnikov, D., et al. (2016) A new optimal transport distance on the space of finite radon measures. *Adv. Differ. Eq.* **11**(12), 1117–1164.
[57] Kondratyev, S. & Vorotnikov, D. (2019) Spherical hellinger–kantorovich gradient flows. *SIAM J. Math. Anal.* **51**(3), 2053–2084.
[58] Kondratyev, S. & Vorotnikov, D. (2020) Convex sobolev inequalities related to unbalanced optimal transport. *J. Differ. Eq.* **268**(7), 3705–3724.
[59] Kondratyev, S. & Vorotnikov, D. (2020) Nonlinear fokker-planck equations with reaction as gradient flows of the free energy. *J. Funct. Anal.* **278**(2), 108310.
[60] Laschos, V. & Mielke, A. (2019) Geometric properties of cones with applications on the hellinger–kantorovich space, and a new distance on the space of probability measures. *J. Funct. Anal.* **276**(11), 3529–3576.
[61] Bihan, D. Le (2014) Diffusion mri: What water tells us about the brain. *EMBO Mol. Med.* **6**(5), 569–573.
[62] Li, B. & Lu, J. (2023) Interpolation between modified logarithmic sobolev and poincare inequalities for quantum markovian dynamics. *J. Stat. Phys.* **190**(10).
[63] Li, B. & Zou, J. (2024) On the convergence of discrete dynamic unbalanced transport models. *ESAIM: Math. Modell. Numer. Anal.* **58**(3), 957–992.
[64] Liero, M., Mielke, A. & Savaré, G. (2016) Optimal transport in competition with reaction: The hellinger–kantorovich distance and geodesic curves. *SIAM J. Math. Anal.* **48**(4), 2869–2911.
[65] Liero, M., Mielke, A. & Savaré, G. (2018) Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. *Invent. Math.* **211**(3), 969–1117.
[66] Lombardi, D. & Maitre, E. (2015) Eulerian models and algorithms for unbalanced optimal transport. *ESAIM: Math. Modell. Numer. Anal.* **49**(6), 1717–1744.
[67] Lombardini, L. & Rossi, F. (2022) Obstructions to extension of wasserstein distances for variable masses. *Proc. Am. Math. Soc.* **150**(11), 4879–4890.
[68] Lott, J. & Villani, C. (2009) Ricci curvature for metric-measure spaces via optimal transport. *Ann. Math.* **169**(3), 903–991.
[69] Maas, J., Rumpf, M., Schönlieb, C. & Simon, S. (2015) A generalized model for optimal transport of images including dissipation and density modulation. *ESAIM: Math. Modell. Numer. Anal.* **49**(6), 1745–1769.
[70] Marchese, A., Massaccesi, A. & Tione, R. (2019) A multimaterial transport problem and its convex relaxation via rectifiable g-currents. *SIAM J. Math. Anal.* **51**(3), 1965–1998.
[71] McCann, R. J. (1997) A convexity principle for interacting gases. *Adv. Math.* **128**(1), 153–179.
[72] Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. In: *Histoire de l'Académie Royale des Sciences de Paris*.
[73] Monsaingeon, L. & Vorotnikov, D. (2021) The schrödinger problem on the non-commutative fisher-rao space. *Calc. Var. Partial Differ. Eq.* **60**(1), 14.
[74] Olkiewicz, R. & Zegarlinski, B. (1999) Hypercontractivity in noncommutative l_p spaces. *J. Funct. Anal.* **161**(1), 246–285.
[75] Otto, F. (2001) The geometry of dissipative evolution equations: The porous medium equation. *Commun. Part. Differ. Eq.* **26**(1-2), 101–174.
[76] Otto, F. & Villani, C. (2000) Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality. *J. Funct. Anal.* **173**(2), 361–400.
[77] Peyré, G., Chizat, L., Vialard, F.-X. & Solomon, J. (2019) Quantum entropic regularization of matrix-valued optimal transport. *Eur. J. Appl. Math.* **30**(6), 1079–1102.
[78] Piccoli, B. & Rossi, F. (2014) Generalized wasserstein distance and its application to transport equations with source. *Arch. Ration. Mech. Anal.* **211**(1), 335–358.
[79] Piccoli, B. & Rossi, F. (2016) On properties of the generalized wasserstein distance. *Arch. Ration. Mech. Anal.* **222**(3), 1339–1365.
[80] Powers, R. T. & Størmer, E. (1970) Free states of the canonical anticommutation relations. *Commun. Math. Phys.* **16**(1), 1–33.
[81] Reid, W. T. (1970) Some elementary properties of proper values and proper vectors of matrix functions. *SIAM J. Appl. Math.* **18**(2), 259–266.
[82] Robertson, J. B., Rosenberg, M., et al. (1968) The decomposition of matrix-valued measures. *Mich. Math. J.* **15**(3), 353–368.
[83] Rockafellar, R. (1971) Integrals which are convex functionals. ii. *Pac. J. Math.* **39**(2), 439–469.
[84] Rouzé, C. & Datta, N. (2019) Concentration of quantum states from quantum functional and transportation cost inequalities. *J. Math. Phys.* **60**(1), 012202.
[85] Rudin, W. (2006) *Real and Complex Analysis*, New York, McGraw-Hill.
[86] Ryu, E. K., Chen, Y., Li, W. & Osher, S. (2018) Vector and matrix optimal mass transport: Theory, algorithm, and applications. *SIAM J. Sci. Comput.* **40**(5), A3675–A3698.

[87] Santambrogio, F. (2015) Optimal transport for applied mathematicians. *Birkäuser, NY* **55**(58-63), 94.
[88] Sturm, K.-T. (2006) On the geometry of metric measure spaces I and II. *Acta Math.* **196**(1), 65–177.
[89] Villani, C. (2003). Topics in optimal transportation. In: *Number 58 in Graduate Studies in Mathematics Number*, American Mathematical Soc.
[90] Villani, C. (2009) *Optimal Transport: Old and New, Volume 338 of Grundlehren der Mathematischen Wissenschaften*, Berlin, Springer-Verlag.
[91] Vorotnikov, D. (2022) Partial differential equations with quadratic nonlinearities viewed as matrix-valued optimal ballistic transport problems. *Arch. Ration. Mech. Anal.* **243**(3), 1653–1698.
[92] Wandell, B. A. (2016) Clarifying human white matter. *Ann. Rev. Neurosci.* **39**(1), 103–128.
[93] Wirth, M. & Zhang, H. (2022) Curvature-dimension conditions for symmetric quantum markov semigroups. *Ann. Henri Poincaré* **24**(3), 1–34.

Appendix A: Auxiliary proofs

Proof of Lemma 4.1. For $\mu \in \mathcal{M}(\mathcal{X}, \mathbb{X})$, by definition, we have $\iota^*_{C(\mathcal{X},\mathcal{O}_\Lambda)}(\mu) = \sup\{\langle \mu, \Xi\rangle_{\mathcal{X}}\,;\, \Xi \in C(\mathcal{X}, \mathcal{O}_\Lambda)\}$. To show that the admissible set $C(\mathcal{X}, \mathcal{O}_\Lambda)$ can be relaxed to $L^\infty_{|\mu|}(\mathcal{X}, \mathcal{O}_\Lambda)$, it suffices to prove

$$\sup_{\Xi \in L^\infty_{|\mu|}(\mathcal{X},\mathcal{O}_\Lambda)} \langle \mu, \Xi\rangle_{\mathcal{X}} \le \sup_{\Xi \in C(\mathcal{X},\mathcal{O}_\Lambda)} \langle \mu, \Xi\rangle_{\mathcal{X}}\,. \tag{A.1}$$

For this, we consider an essentially bounded measurable field $\Xi \in L^\infty_{|\mu|}(\mathcal{X}, \mathcal{O}_\Lambda)$. Without loss of generality, we assume that it is bounded by $\|\Xi\|_\infty$ everywhere. By Lusin's theorem, for any $\varepsilon > 0$, there exists a continuous field with compact support $\widetilde{\Xi}$ such that

$$|\mu|(\{x \in \mathcal{X};\, \Xi(x) \neq \widetilde{\Xi}(x)\}) \le \varepsilon\,. \tag{A.2}$$

Define $\mathbb{P}_{\mathcal{O}_\Lambda}$ as the L^2-projection from $\mathbb{X}$ to the closed convex set $\mathcal{O}_\Lambda$. By abuse of notation, we still denote by $\widetilde{\Xi}$ the composite function $\mathbb{P}_{\mathcal{O}_\Lambda} \circ \widetilde{\Xi} \in C(\mathcal{X}, \mathcal{O}_\Lambda)$. It is clear that $\|\widetilde{\Xi}\|_\infty \le \|\Xi\|_\infty$, and (A.2) still holds. Then it follows that $|\langle \mu, \Xi\rangle_{\mathcal{X}} - \langle \mu, \widetilde{\Xi}\rangle_{\mathcal{X}}| \le 2\varepsilon \|\Xi\|_\infty$, which further implies

$$\langle \mu, \Xi\rangle_{\mathcal{X}} \le \langle \mu, \widetilde{\Xi}\rangle_{\mathcal{X}} + 2\varepsilon\|\Xi\|_\infty \le \sup_{\Xi \in C(\mathcal{X},\mathcal{O}_\Lambda)} \langle \mu, \Xi\rangle_{\mathcal{X}} + 2\varepsilon\|\Xi\|_\infty\,.$$

Since ε is arbitrary, we have proved the claim (A.1). Thus, we can take the pointwise sup in (4.4) and obtain the desired $\iota^*_{C(\mathcal{X},\mathcal{O}_\Lambda)}(\mu) = \mathcal{J}_{\Lambda,\mathcal{X}}(\mu)$ by Proposition 3.1. Next, we characterise the subgradient $\partial \mathcal{J}_{\Lambda,\mathcal{X}}(\mu)$. By Lemma 2.4, we have $\Xi \in \partial \mathcal{J}_{\Lambda,\mathcal{X}}(\mu) \bigcap C(\mathcal{X}, \mathbb{X})$ if and only if $\langle \mu, \Xi\rangle_{\mathcal{X}} = \iota_{C(\mathcal{X},\mathcal{O}_\Lambda)}(\Xi) + \mathcal{J}_{\Lambda,\mathcal{X}}(\mu)$, which yields $\Xi \in C(\mathcal{X}, \mathcal{O}_\Lambda)$ and

$$\int_{\mathcal{X}} \mu_\lambda \cdot \Xi - J_\Lambda(\mu_\lambda)\, \mathrm{d}\lambda = 0\,, \tag{A.3}$$

where λ is a reference measure such that $|\mu| \ll \lambda$ and μ_λ is the density of μ. We note from $J_\Lambda = \iota^*_{\mathcal{O}_\Lambda}$ and $\Xi(x) \in \mathcal{O}_\Lambda$ that $\mu_\lambda \cdot \Xi - J_\Lambda(\mu_\lambda) \le 0$, λ-a.e., where by (A.3), the equality actually holds λ-a.e.. Then (4.5) follows. □

Proof of Lemma 5.1. It suffices to consider $[a, b] = [0, 1]$. We denote by $\widetilde{\mathrm{WB}}_\Lambda$ the right-hand side of (5.1). By Hölder's inequality and recalling ($\mathcal{P}$) with the admissible set $\widetilde{\mathcal{CE}}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$, we have $\widetilde{\mathrm{WB}}_\Lambda \le \mathrm{WB}_\Lambda$. For the other direction, we consider $\{\mu_t\}_{t\in[0,1]} \in \widetilde{\mathcal{CE}}([0, 1]; \mathsf{G}_0, \mathsf{G}_1)$ and reparameterize it by the ε-arc length function $s = \mathsf{s}_\varepsilon(t)$:

$$s = \mathsf{s}_\varepsilon(t) = \int_0^t \left(\mathcal{J}_{\Lambda,\Omega}(\mu_\tau)^{1/2} + \varepsilon\right) \mathrm{d}\tau : [0, 1] \to [0, L(\mu_t) + \varepsilon]\,,$$

where $L(\mu_t) := \int_0^1 \mathcal{J}_{\Lambda,\Omega}(\mu_\tau)^{1/2}\, \mathrm{d}\tau$. It is clear that $\mathsf{s}_\varepsilon(t)$ is strictly increasing and absolutely continuous and has an absolutely continuous inverse. Then, by Lemma 3.15 and writing $\widetilde{\mu}^\varepsilon_s = \mu_{\mathsf{s}_\varepsilon^{-1}(s)}$ for short, we have

$$\mathrm{WB}^2_\Lambda(\mathsf{G}_0, \mathsf{G}_1) \le (L(\mu_t) + \varepsilon) \int_0^{L(\mu_t)+\varepsilon} \mathcal{J}_{\Lambda,\Omega}(\widetilde{\mu}^\varepsilon_s)\, \mathrm{d}s = (L(\mu_t) + \varepsilon) \int_0^1 \frac{\mathcal{J}_{\Lambda,\Omega}(\mu_t)}{\mathcal{J}_{\Lambda,\Omega}(\mu_t)^{1/2} + \varepsilon}\, \mathrm{d}t\,, \tag{A.4}$$

https://doi.org/10.1017/S0956792524000901 Published online by Cambridge University Press

where the first inequality is by ($\mathcal{P}'$) with $[a,b]=[0,L(\mu_t)+\varepsilon]$. Letting $\varepsilon\to 0$ in (A.4), we can find $\mathrm{WB}_\Lambda \le \widetilde{\mathrm{WB}}_\Lambda$. If we assume that μ minimises ($\mathcal{P}$), we have

$$\mathrm{WB}_\Lambda(\mathsf{G}_0,\mathsf{G}_1)=\Big(\int_0^1 \mathcal{J}_{\Lambda,\Omega}(\mu_t)\,\mathrm{d}t\Big)^{1/2}\le\int_0^1 \mathcal{J}_{\Lambda,\Omega}(\mu_t)^{1/2}\,\mathrm{d}t\,,$$

which implies that $\mathcal{J}_{\Lambda,\Omega}(\mu_t)$ is constant a.e.. Then (5.2) immediately follows. □

Cite this article: Li B. and Zou J. On a general matrix-valued unbalanced optimal transport problem. *European Journal of Applied Mathematics*, https://doi.org/10.1017/S0956792524000901